(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 684 312 A2**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **95105973.2**

(22) Date of filing: **21.04.95**

(51) Int. Cl.[6]: **C12N 15/29**, C07K 14/43, C12N 15/62, C12N 1/15, C12N 15/80, //(C12N1/15, C12R1:66,1:80)

The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(30) Priority: **21.04.94 ES 9400836**

(43) Date of publication of application: **29.11.95 Bulletin 95/48**

(84) Designated Contracting States: **AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE**

(71) Applicant: **URQUIMA S.A.**
Decano Bahi, 67
E-08026 Barcelona (ES)

(72) Inventor: **Uriach-Marsal, Juan**
Dr. Fleming 18-20
E-08017 Barcelona (ES)
Inventor: **Rubio-Susan, Victor**
Cuesta San Vincente 20
E-28008 Madrid (ES)
Inventor: **Patino-Martin, Cristina**
Sector Literatos 32
E-28760 Tres Catos,
Madrid (ES)
Inventor: **Iossif Kalo-Koenova, Eliza**
Marqués de Zafra 28
E-28028 Madrid (ES)
Inventor: **del-Moral-Juàrez, Catalina**
Avda. Santa Eugenia 58
E-28031 Madrid (ES)
Inventor: **Faus-Santasusana, Ignacio**
Arquitecto Moragas 18A
E-08035 Barcelona (ES)
Inventor: **del-Rio-Pericacho, José-Luis**
Rambla d'Egara 273
E-08224 Tarrasa,
Barcelona (ES)
Inventor: **Bladé-Piqué, Joan**
Albert Llanas 11
E-08024 Barcelona (ES)

(74) Representative: **Zumstein, Fritz, Dr. et al**
Dr. F. Zumstein
Dipl.-Ing. F. Klingseisen
Bräuhausstrasse 4
D-80331 München (DE)

(54) Preparation process of a natural protein sweetener.

(57) Thaumatin II or thaumatin I can be obtained through the expression, not of their natural genes, but of artificial, synthetic and substantially optimized genes following specific rules. Preferably, this expression is carried out in filamentous fungi, especially GRAS fungi and particularly the species Penicillium roquefortii, Aspergillus niger and the awamori variant of Aspergillus niger. Preparing substantially optimized artificial genes for filamentous fungi, performed here for the first time in the case of thaumatin, allows for high protein expression, making the process useful for industrial production of this valuable sweetener. Thaumatins may be obtained extracellularly by using a plasmid with a secretion signal, and also intracellularly. The latter method can be used in animal feed without prior separation from the fungal mycelium.

EP 0 684 312 A2

This invention is based on genetic engineering or recombinant DNA technology and refers to a process for obtaining natural proteinaceous sweeteners of the thaumatin type, to new DNA sequences which have been optimized for expression in filamentous fungi and which codify these proteins, and to the use of these sequences in the transformation of filamentous fungi for the production of thaumatins.

STATE OF THE ART

The thaumatins are proteins with a very sweet taste and the capacity to increase the palatability (upgrading or improving other flavours) of food; in industry they are currently extracted from the arils of the fruit of the plant Thaumatoccocus daniellii Benth. Thaumatins can be isolated from these arils in at least five different forms (I, II, III, b and c), which can be separated using ion-exchange chromatography. These forms are all single-chain polypeptides with 207 amino acids and a molecular weight of approximately 22,000 Daltons. Thaumatins I and II, which predominate in the arils and have very similar sequences of amino acids, are much sweeter than saccharose (100,000 times sweeter according to one estimate). Besides being natural products, thaumatins I and II are non-toxic, making them a good substitute for common sweeteners in the animal and human food industries.

Despite its advantages, industrial use of thaumatins of natural plant origin is very limited because of the extreme difficulty involved in obtaining the fruit from which it is extracted. The producing plant, T. daniellii, not only requires a tropical climate and pollination by insects, but it must also be cultivated among other trees and yet 75% of its flowers do not bear fruit.

Although attempts have been made to produce thaumatins by genetic engineering in bacteria such as Escherichia coli (cf. EP 54.330, EP 54.331 and WO 89/06283), Bacillus subtilis and Streptomyces lividans, in yeasts such as Saccharomyces cerevisiae (cf. WO 87/03007) and Kluveromyces lactis (EP 96.430 and EP 96.910), in the fungus Aspergillus oryzae (Hahm and Batt, Agric. Biol. Chem. 1990, vol. 54, pp. 2513-20), and in transgenic plants such as Solanum tuberosum, until now the results have been considered disheartening; thus the thaumatin available to industry is very scarce and expensive (cf. M. Witty and W.J. Harvey, "Sensory evaluation of transgenic Solanum tuberosum producing r-thaumatin II", New Zealand Journal of Crop and Horticultural Science, 1990, vol. 18, pp. 77-80, and the articles cited therein).

Accordingly, there has remained a need for economically obtaining industrial amounts of thaumatins.

DESCRIPTION OF THE INVENTION

This invention solves the problem of preparing thaumatins II and I through their expression in filamentous fungi but without using natural DNA (or derived cDNA) as described for the fungus Aspergillus oryzae. Rather, artificial, synthetic and substantially optimized genes are used for expression in filamentous fungi according to specific rules.

Obtaining substantially optimized artificial genes for filamentous fungi, performed here for the first time for thaumatins, allows for high expressions of protein, making the process useful for industry.

In a specific embodiment of this invention, the filamentous fungi used belong to those considered innocuous, particularly to those included on the GRAS list (Generally Recognized as Safe). Preferred GRAS fungi include the Penicillium genus, especially the species Penicillium roquefortii, or the Aspergillus genus, especially the niger species and the niger variant awamori.

This invention encompasses obtaining thaumatins I and II secreted or produced extracellularly (for which an appropriate secretion signal must be introduced in the plasmid), and obtaining thaumatins I and II intracellularly, which allows for their use in animal food, without prior separation of the mycelium from the fungi.

The following abbreviations are used below, among others:

A	= Adenine
Amp	= Ampicillin
ATP	= Adenosine triphosphate
BSA	= Bovine serum albumin
C	= Cytosine
CIP	= Calf intestinal phosphatase
dATP	= 2'-Deoxyadenosine triphosphate
dCTP	= 2'-Deoxycytidine triphosphate
dGTP	= 2'-Deoxyguanosine triphosphate
DNA	= deoxyribonucleic acid
DTT	= 1,4-Dithiothreitol

dTTP = 2'-Deoxythymidine triphosphate
EDTA = Ethylenediaminetetra-acetic acid (disodium salt)
G = Guanine
GRAS = Generally regarded as safe
KDa = Kilodalton
MCS = Multiple cloning site
nt = Nucleotides
bp = base pairs
PCR = Polymerase chain reaction
PEG = Polyethylene glycol
PMSF = Phenylmethylsulfonyl fluoride
rpm = revolutions per minute
SDS = Sodium dodecyl sulphate
SSC = Sodium sodium citrate (0.15M NaCl; 0.015M sodium citrate)
T = Thymine
TE = Buffer 10 mM Tris-HCl, pH 8.0; 1 mM EDTA
U = Units
X-gal = 5-bromo-4-chloro-3-indo-β-D-galactose

Amino acids are designated by their standard abbreviations. For plasmids, the published notation in each case is used.

One part of the subject-matter of this invention is a gene for codifying thaumatin II which is artificial, synthetic and more than 50% optimized for expression in filamentous fungi; this gene consists of a DNA sequence which codifies the sequence of amino acids of Sequence ID No. 1 (corresponding to the 207 amino acids of the protein thaumatin II), followed by n stop sequences, where integer n is greater than or equal to 1; this DNA sequence is the result of making more than 50% of the possible modifications of the DNA sequence of the natural gene which codifies the 207 amino acids of thaumatin II (gene described in the literature and also included in Sequence ID No. 1) through the addition of one or more (n in Sequence ID No. 1) stop codons and performing more than 50% of the possible changes on the nucleotide codons corresponding to the thaumatin II amino acids; these changes consist of substituting the original codons in a given amino acid with the codon in parentheses in the following list of amino acid codons:
Ala (GCC), Arg (CGC), Asn (AAC), Asp (GAC), Cys (TGC), Lys (AAG), Gln (CAG), Glu (GAG), Gly (GGC), Ile (ATC), Leu (CTC), Met (ATG), Phe (TTC), Pro (CCC), Ser (TCC), Thr (ACC), Trp (TGG), Tyr (TAC), Val (GTC);
As is well known in the art, TAA, TAG or TGA can be used as stop codons, or any combination thereof.

The specific case of the previous gene in which an optimization of more than 75% was performed is preferred. It is even more preferred when the optimization is maximum (100%), i.e., when the DNA sequence of the artificial gene is obtained from the Sequence ID No. 1 sequence by performing 100% of the all possible codon changes, which corresponds to Sequence ID No. 2. Also preferred are the previous genes where n is between 1 and 3.

Another part of the subject-matter of this invention is a gene for codifying thaumatin I which is artificial, synthetic and more than 50% optimized for its expression in filamentous fungi; this gene consists of a DNA sequence which codifies the sequence of amino acids corresponding to the 207 amino acids of the protein thaumatin I (sequence of 207 amino acids which differs from those of Sequence ID No. 1 in only five amino acids, i.e., 46-Asn, 63-Ser, 67-Lys, 76-Arg and 113-Asn); this optimized DNA sequence is obtained by leaving the following five codons unchanged: AAC (46-Asn), TCC (63-Ser), AAG (67-Lys), CGC (76-Arg)and AAC (113-Asn), by modifying the rest of the codons as described above for the DNA sequence of the thaumatin II gene, and by adding one or more stop codons, as described above. The gene which codifies thaumatin I and which is more than 75% optimized is particularly preferred. It is even more preferred when the optimization is maximum (100%). Artificial genes to which between one and three stop codons have been added are preferred.

Hereinafter, any gene optimized more than 50%, more than 75% or up to 100% is called without distinction a "substantially optimized gene".

Subject-matter of this invention are also the recombinant plasmids made up of: (i) a substantially optimized gene for obtaining thaumatin I or II, (ii) an expression cassette for filamentous fungi containing an appropriate promoter sequence and a terminating sequence for this type of fungi, (iii) an appropriate selection marker, and (iv) an optional secretion signal DNA sequence for producing the protein extracellularly.

Particularly preferred are recombinant plasmids characterized in that the promoter sequence of the expression cassette comes from the gene of the enzyme glyceraldehyde 3-phosphate dehydrogenase of Aspergillus nidulans; the terminating sequence of the expression cassette is the tryptophan C sequence of Aspergillus nidulans; and the selection marker is that of resistance to sulfanilamide. Also preferred are the recombinant analogue plasmids where the promoter sequence of the expression cassette comes from the gene of the enzyme glucoamylase of Aspergillus niger.

In a particular embodiment of this invention, the recombinant plasmids used express the fusion protein thaumatin-glucoamylase, and they are characterized by comprising: (i) an appropriate selection marker; (ii) a DNA sequence made up of (a) a substantially optimized gene for the expression of thaumatin I or II, (b) a spacer sequence which in turn contains a KEX2 processing sequence, and (c) the complete gene of the glucoamylase of Aspergillus niger or the awamori (glaA) variant of Aspergillus niger; and (iii) the "pre" and "pro" signal sequences of the glaA gene.

Part of the subject-matter of this invention are also the cultures of filamentous fungi capable of producing the proteins thaumatin I or II, which have been transformed with any of the abovementioned plasmids. In particular, the filamentous fungi of the species Penicillium roquefortii, Aspergillus niger and the awamori variant of Aspergillus niger are preferred.

Part of the subject-matter of this invention are also the production processes for thaumatin I or II which include the following steps:

a) incorporation of a substantially optimized gene for the expression of thaumatin I or II, in an expression vector selected from those corresponding to the abovementioned plasmids using standard recombinant DNA technology techniques;

b) transformation of a strain of filamentous fungus with the previous expression vector;

c) culture of a filamentous fungus strain transformed in this way in the appropriate nutrient conditions to produce thaumatin I or II, either intracellularly, extracellularly or through both methods simultaneously, or in the form of the fusion protein thaumatin-glucoamylase;

d) depending on the case, separation and purification of thaumatin I or II alone, or separation of thaumatin I or II from the culture medium, together with the fungal mycelium.

In a preferred embodiment of these processes, the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger or the awamori variant of Aspergillus niger.

To obtain thaumatin II, pThII recombinant plasmids are preferred, which can be obtained through the method described in the examples and illustrated in Figure 6, which can be summarized as follows: a) starting with plasmid pTZ18RN(3/4), a fragment (3/4) of the DNA sequence of the substantially optimized gene which codifies thaumatin II is obtained; b) this fragment is ligated with plasmid pAN52-3, generating plasmid pTh(3/4); c) starting with plasmid pTZ18RN(1/2), the remaining fragment (1/2) of the DNA sequence of the substantially optimized gene which codifies thaumatin II is obtained; d) this fragment is ligated to plasmid pTh(3/4), generating plasmid pTh; e) a DNA fragment is inserted to provide resistance to sulfanilamide, Su^r, thus obtaining plasmid pThII (Figure 6). With this plasmid, thaumatin II is obtained intracellulary for the most part.

For the production of thaumatin II in a basically extracellular way in Penicillium roquefortii, pThIII plasmids are preferred, the preparation of which is described in Example 2 and is outlined in Figure 9. To prepare it in the awamori variant of Aspergillus niger, the process described in Example 3 is used.

To produce thaumatin II as a fusion protein with glucoamylase, the pECThII and pThIX plasmids can be used, preparation of which is described in the examples and outlined in Figures 12, 13A and 13B.

To produce thaumatin I, the recombinant plasmids obtained following methods analogous to those used to produce thaumatin II are used. Thus, for example, for intracellular production in Penicillium roquefortii, pThI plasmids are used which are obtained as follows: a) Starting with plasmid pTZ18RN(1/2), the fragment (1/2) of the substantially optimized gene sequence is obtained which codifies thaumatin II; b) this fragment is ligated to plasmid pTZ18RN(3/4) linearized with NcoI, thus generating plasmid PTZ18RN(Th); c) starting with plasmid pTZ18RN(Th) in single-stranded form and using site-directed mutagenesis techniques, the following changes are carried out on the sequence of the synthetic and artificial gene of thaumatin II, where the symbol -> joins the replaced (original) and the replacement (final) in this order:

AAG -> AAC (46-Lys -> 46-Asn)
CGC -> TCC (63-Arg -> 63-Ser)
CGC -> AAG (67-Arg -> 67-Lys)
CAG -> CGC (76-Gln -> 76-Arg)
GAC -> AAC (113-Asp -> 113-Asn)

This plasmid is called pTZ18RN(ThI); d) starting with plasmid PTZ18RN(ThI) a DNA fragment of the complete sequence of the substantially optimized gene which codifies thaumatin I is obtained; e) this

fragment is ligated to plasmid pAN52-3, thus generating plasmid pTh'; f) a DNA fragment containing resistance to sulfanilamide, Su^R, is inserted, thus obtaining plasmid pThI.

In a specific embodiment of this invention, the plasmids are replicated and amplified in Escherichia coli.

When the filamentous fungus is of the GRAS type, the processes for isolating thaumatin I or II together with the fungal mycelium are particularly interesting. In these cases, a part of the subject-matter of this invention is also the use of mixtures of thaumatin I or II and fungal mycelium obtained in this way to increase the sweetness or palatability of animal food.

When it is necessary to obtain purified thaumatin I or II, it is particularly important for the expression vector to be a plasmid which also contains a secretion signal sequence in the DNA so that the filamentous fungus produces thaumatin I or II extracellularly.

In some cases the production of thaumatin I or II can be increased by obtaining the fusion protein with glucoamylase.

In specific embodiments of this invention, when obtaining the pThI and pThII plasmids, the promoter sequence of the expression cassette can come from any gene from the following enzymes of filamentous fungi: glyceraldehyde 3-phosphate dehydrogenase, β-glucoamylase, alcohol dehydrogenase, glucoamylase or α-amylase. Moreover, the terminating sequence of the expression cassette can be the sequence corresponding to the promoter sequence in question. Finally, the selection marker can be of the type which is resistent to sulfanilamide, oleomycin, hygromycin B, phleomycin or acetamide.

As shown in the examples, this invention makes it possible to obtain thaumatin I or II for industry with satisfactory phenotypical characteristics, and with high productivity, which represents a considerable advantage over the state of the art.

Moreover, because the fungus is harmless, the thaumatin can be administered together with the mycelium, a fact which saves time in the purification process and, therefore, represents a considerable additional advantage, especially for use in animal feed.

Without being limiting, the following detailed examples illustrate this invention. The culture of the fungus Penicillium roquefortii, which produces the thaumatin II obtained in Example 1, has been deposited in the Spanish Collection of Standard Cultures (Colección Española de Cultivos Tipo, CECT) of the Departamento de Microbiología of the Facultad de Ciencias Biológicas of the University of Valencia, with number CECT 2972.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: (A) DNA sequence showing nucleotides 272-304 from the MCS of commercial plasmid pTZ18R. (B) Fragment of plasmid pTZ18RN, obtained from the former, showing its unique NcoI restriction site.

Figure 2: Strategy used to build the synthetic gene with two pairs of oligonucleotides. Each pair of oligonucleotides has a complementary area. A, B and C represent restriction enzymes necessary for cloning of the oligonucleotide pairs, once they are paired and elongated, on the pTZ18RN vector.

Figure 3: Sequences of the oligonucleotides used to build the gene.

Figure 4: Diagram of the different stages in the construction of the artificial and synthetic gene (sequence represented in black).

Figure 5: Representative autoradiographs of the gene sequence using the Sanger dideoxy method: (A) the first 60 nucleotides; (B) nucleotides 70-170; (C) nucleotides 330-370.

Figure 6: Diagram of the manipulations performed to obtain the pThII plasmid.

Figure 7: Results of the PCR analysis of the two transformed fungi, M0901 and T0901, compared with the pThII plasmid and an untransformed control fungus. On the y-axis, the number of bases according to two standard reference markers.

Figure 8: Results of the immunoblotting analysis of the transformed fungi from Figure 7, compared with commercial thaumatin II (supplied by Sigma Inc.) and an untransformed control fungus (E = extracellular protein; I = intracellular protein). The numbers on the y-axis correspond to protein markers of known molecular weight. The arrow indicates the place where the comercial thaumatin (4) and the recombinant thaumatin migrate (2, 3, 5 and 6).

Figure 9: Diagram of the manipulations performed to obtain plasmid pThIII. The sequence corresponding to the gene of resistance to sulfanilamide (Su^R) is shown as the dark crosshatched section and the sequence of thaumatin is shows as the lighter crosshatched section. The section with vertical lines shows the different fungal promoter and terminating sequences, as well as the "signal" sequence of 24 amino acids from the glucoamylase gene (labelled $SSGlaA_{24}$ in the figure).

Figure 10: Results of PCR analysis of the A2 transformed fungus (thaumatin secretor). On the x-axis, the number of bases according to standard reference markers. Lanes 1 and 5 correspond to markers, lane 2 contains DNA from an untransformed fungus (control), and lane 3 contains DNA from fungus a2. Lane 4 is a positive control (DNA from plasmid pThIII).

Figure 11: Results of the immunoblotting analysis of the transformed fungi T0901 and a2. Lane 1 contains commercial thaumatin supplied by Sigma, Inc. Lane 7 corresponds to protein markers of known molecular weight (the molecular weights of each protein are indicated next to each lane). Lane 2 contains the culture medium in which the T09011 fungus was grown, a producer of intracellular thaumatin. Lanes 3 and 4 contain the culture medium in which the a2 fungus was grown (extracellular producer) and an untransformed fungus (control). Lanes 5 and 6 contain mycelium from these two fungi, respectively.

Figure 12: Diagram of the manipulations performed to obtain the pECThII plasmid. The dark cross-shatched section represents the synthetic gene of tharnmatin II.

Figures 13A and 13B: Diagram of the manipulations performed to obtain the pThIX plasmid. The dark crosshatched section is the glucoamylase (glaA) sequence of Aspergillus niger or the awamori variant of Aspergillus niger. The wavy line section represents the glutathione-S-transferase sequence of Escherichia coli. The synthetic gene codifying thaumatin II appears as the lighter grey crosshatched section and the spacer sequence is between the genes of thaumatin and glucoamylase with vertical lines.

Figure 14: Details of the sequences in the fusion area between glucoamylase and thaumatin.

EXAMPLES

EXAMPLE I: INTRACELLULAR PRODUCTION OF THAUMATIN II IN PENICILLIUM ROQUEFORTII

(1.1) Construction of the synthetic, artificial and completely optimized gene encoding thaumatin II.

(1.1.1) Optimization of the DNA sequence of thaumatin II.

Starting with the sequences of known amino acids and nucleotides in the bibliography for thaumatin II and its corresponding natural gene (cf. for example: EP 54.330), reproduced in Sequence ID No. 1, the sequence of optimized DNA of Sequence ID No. 2 was designed, which codifies the same protein and where n = 3 (it has 3 TAA stop codons). The optimized sequence of Sequence ID No. 2 was obtained by performing the maximum number of changes on the codons of Sequence ID No. 1, replacing the original codons with the codons indicated in parenthesis on the following list of amino acid codons, when the latter where different from the originals:

Ala (GCC), Arg (CGC), Asn (AAC), Asp (GAC), Cys (TGC), Lys (AAG), Gln (CAG), Glu (GAG), Gly (GGC), Ile (ATC), Leu (CTC), Met (ATG), Phe (TTC), Pro (CCC), Ser (TCC), Thr (ACC), Trp (TGG), Tyr (TAC), Val (GTC);

(1.1.2) Construction of the pTZ18RN recombinant plasmid using site-directed mutagenesis.

Before beginning assembly of the synthetic gene of thaumatin II, a single NcoI restriction site was inserted in the multiple cloning site (MCS) of the multifunctional plasmid pTZ18R (supplied by Pharmacia Inc.). In this way plasmid pTZ18RN was generated ("N" for NcoI), the restriction site of which is shown in Figure 1. The insertion of the NcoI restriction site was performed using the site-directed mutagenesis technique described below:

Oligonucleotide p115 (5'-ACCCGGGGATCCTCTCCATGGGACCTGCAGGCATGCA-3') was supplied by Ingenasa S.A. (Madrid, Spain). Using standard procedures (Maniatis et al., "Molecular cloning, a laboratory manual", Cold Spring Harbor Laboratory Press, 1989), this oligonucleotide was labeled at the 5, end by transferring ^{32}P from [gamma-^{32}P]ATP with polynucleotide kinase. pTZ18R. with its DNA in single-stranded form, was obtained by standard techniques and was hybridized with one picomol of oligonucleotide labelled with ^{32}P at the 5, end in a buffer containing 40 mM Tris.HCl, pH 7.5, 50 mM NaCl and 20 mM $MgCl_2$ (final volume 5 µL). The mixture was incubated at 65 °C for five minutes and allowed to cool slowly (overnight) to room temperature. The following enzymes and reagents were then added to the 5 µL of this mixture: 1.5 µL of 10X solution B (200 mM Tris.HCl, pH 7.5; 100 mM $MgCl_2$; 50 mM DTT); 1 µL of 10 mM ATP; 4 µL of a mixture containing 2.5 mM of each of the 4 dNTPs (dATP, dGTP, dTTP, dCTP); 6.5 µL of water; 1 µL of T4 DNA polymerase (3 units/µL); and 1 µL of DNA ligase (6 units/µL). The reactions were incubated for 3 hours at room temperature and at the end of that time 1 µL of T4 DNA polymerase was added (3 units) and 1 µL of DNA ligase (6 units). The reactions were allowed to continue for 60 more minutes at 37 °C.

Aliquots of 1.0 µL of each reaction were used to transform E. coli strain JM103. Various colonies grown in LB/ampicillin (100 µg/mL) dishes were replated in dishes with fresh medium and analyzed (LB = Luria broth, a culture medium with the following composition: 1% bacto-tryptone, 0.05% yeast extract, 170 mM NaCl, pH 7.0). To be able to identify the clones containing the desired mutation, the colonies were analyzed using the p115 oligonucleotide labelled with [gamma-^{32}P]ATP as a probe, as described below.

Candidate colonies were replated in nitrocellulose filters (Schleicher & Schuell). The filters were placed in LB/amp dishes and incubated overnight at 37 °C. The next day the cells were lysed by successively washing the filters in three solutions:

- Five minutes in 0.5 M Tris.HCl, pH 7.5, 1 M NaCl.
- Five minutes in 1 M Tris.HCl, pH 7.5.
- Five minutes in 0.5 M Tris.HCl, pH 7.5, 1 M NaCl.

The filters were then dried at 80 °C for 90 minutes. Once they were dry, the filters were washed three times in 3X SSC, 0.1% SDS. Pre-hybridization took place in a solution containing 6X SSC, 5X Denhardt solution, 0.05% sodium pyrophosphate, 100 µg/ml of boiled salmon sperm DNA, and 0.5% SDS. Filters were pre-hybridized for one hour at 37 °C. Hybridization took place overnight in 50 mL of the same solution, to which 33 ng of labelled p115 probe was added. The hybridization temperature was 50 °C. On the next day the filters were washed as follows:

- First wash: 15 minutes in 2X SSC, 0.1% SDS, at room temperature.
- Second wash: the same conditions, but at 55 °C.
- Third wash: The same conditions, but at 65 °C.
- Fourth wash: 15 minutes in 0.4X SSC, 0.1% SDS at 65 °C.

After the fourth wash, the filters were exposed to an X-ray film for 2 hours at -20 °C. Various colonies with DNA showing marked hybridization to probe 115 were identified and DNA was extracted from each one.

The final identity of the clones was verified by testing if the DNA could be cut or not cut with NcoI and by analyzing its sequence. The plasmid containing the NcoI restriction site between the BamHI and PstI restriction sites (Figure 1) was called pTZ18RN and was the parent vector used in the construction of the artificial, synthetic and totally optimized gene of thaumatin II.

(1.1.3) Strategy for building the synthetic gene which codifies thaumatin II

The method chosen for assembling the synthetic gene of thaumatin II is shown in Figure 2. The eight long oligonucleotides whose sequences are shown in Figure 3 were supplied by Isogen Bioscience, Inc. (Netherlands). The single-stranded oligonucleotides, which occur in pairs, can be paired because of the complementary nature of the sequences. They were labelled 1a, 1b,; 2a, 2b; 3a, 3b; and 4a, 4b. After pairing, the single-stranded areas were filled with the modified T7 DNA polymerase (the Taq DNA polymerase can also be used). The resulting double-stranded fragments were digested with the appropriate restriction enzymes to obtain cohesive ends or blunt ends and then ligated to the desired vector.

Figure 4 shows the strategy used to build the synthetic gene in 2 fragments which were then joined to an expression vector.

(1.1.3.1) Assembly of the first 332 pairs of bases of the synthetic gene of ID Sequence No. 2 (n = 3).

In the first stage, the oligonucleotides 1a, 1b, 2a and 2b were joined to obtain a DNA fragment with 332 base pairs which could be inserted into the pTZ18RN plasmid.

One microgram of oligonucleotide 1a and 1 µg of 1b were mixed in a buffer solution containing 40 mM Tris.HCl, pH 8.0, 10mM $MgCl_2$, 5mM DTT, 50 mM NaCl and 50 µg/mL of bovine serum albumin (BSA). The mixture (17 µL) was heated for 5 minutes at 70 °C and then cooled slowly to 65 °C for about ten minutes (appropriate temperature for hybridizing the pairs of oligonucleotides). Then 2 µL of a mixture of the four deoxynucleotides was added (2.5 mM of each dNTP) and 1 µL of the modified T7 DNA polymerase enzyme (Sequenase brand from U.S. Biochemical Corp.), giving a final volume of 20 µL. The reactions took place for 30 minutes at 37 °C, followed by 10 additional minutes at 70 °C (to inactivate the Sequenase). The reaction products were digested with Bam HI and Bgl II at 37 °C for 3 hours. The following extractions were performed on the DNAs: once with phenol, once with phenol:chloroform and once with chloroform; they were then precipitated with ethanol. They were finally frozen in TE buffer at -20 °C until later use.

The 2a and 2b oligonucleotides were processed in the same way except that the final products were digested with Bgl II and Nco I.

Plasmid pTZ18RN was digested sequentially with Bam HI and Nco I and was dephosphorylated with calf intestinal phosphatase (CIP). The linearized fragment of 2871 pairs of bases was recovered from a 0.8% agarose gel and then purified.

Then the products of reactions 1 and 2 were joined with the linearized pTZ18RN and the mixture was used to transform E. coli strain NM522. To identify the clones with the insert, a white/blue indicator test was used which works basically as follows:

The pTZ18R plasmid and its derivative pTZ18RN contain the bacterial gene LacZ'. Therefore, the bacterial colonies containing this plasmid are blue on dishes with LB/ampicillin which also contain the chromogenic substrate 5-Bromo-4-chloro-3-indo-β-D-galactose (X-gal). When a fragment of foreign DNA is inserted in the multiple cloning site (MCS) of the pTZ18RN plasmid, the LacZ' gene is deactivated and the resulting colonies are not blue. but white. Therefore, the white colonies were initially isolated, given that they were candidates for containing the different fragments of the synthetic gene of thaumatin II.

Various colonies with inserts of the appropriate size contained complete fragments of the 325 base pairs of the synthetic gene of thaumatin II. The resulting plasmid was called pTZ18RN(1/2).

(1.1.3.2) Assembly of the second 305 pairs of bases of the synthetic gene of ID Sequence No. 2 (n = 3)

In this case, an alternative approach was put into practice using Taq DNA polymerase and the PCR technique.

Before the annealing stage, oligonucleotides 3b and 4a were labelled at their 5' ends with a phosphate group using standard techniques. The oligonucleotides were called 3b* and 4a*.

One microgram of 3a and 1 μg of 3b* were incubated in a reaction mix (18 μL) containing 10 mM Tris.HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.1 mg/ml of gelatin. The samples were incubated for 5 minutes at 70 °C and for five more minutes at 65 °C. At this point, each dNTP was added (G, A, T, C) at a final concentration of 2 mM and 2.5 units of Ampli Taq DNA polymerase (Perkin-Elmer Cetus). The PCRs were as follows: 1 minute at 94 °C; 1 minute at 55 °C; and 1 minute at 72 °C for 30 cycles, followed by a final extension at 72 °C for 5 minutes. The samples were then extracted with phenol:chloroform and resuspended in 10 μL of TE buffer and incubated with Nco I at 37 °C for 3 hours. After extracting and precipitating with ethanol, the DNAs were dissolved in TE buffer and frozen at -20 °C until later use.

The 4a* and 4b oligonucleotides were processed as described above, except that the final products were digested with Pst I.

Ligation of the three fragments was done as per the same process mentioned above, except that pTZ18RN was used which was cut with Nco I and Pst I, treated with calf intestinal phosphatase and finally purified from an agarose gel. The ligation reactions contained 15% polyethylene glycol (PEG), which stimulates ligations with blunt ends. The ligation products are used to transform E. coli NM 522. A white/blue selection was made again of the recombinants on dishes with LB/amp medium supplemented with X-gal and IPTG. After analyzing the transformants, one clone was isolated which contained the 305 pb fragment of the second part of the thaumatin II gene. This plasmid was called pTZ18RN (3/4).

(1.1.3.3) Sequence Analysis

The identity of the synthetic gene was verified by analyzing its sequence using the Sanger method (Sanger, F. et al., Proc. Nat. Acad. Sci. USA 1977, vol. 74, p. 5463-67). A sequentiation kit was used (version 2.0) from United States Biochemical Corp. The sequence of the synthetic gene was determined without ambiguity by: (1) sequentiation of the two gene strands; and (2) performing parallel sequentiation reactions with dITP to destabilize the potential secondary structures which could form due to the areas rich in GC. Representative autoradiographs are shown in Figure 5.

(1.2) Insertion of the gene in an expression vector for filamentous fungi (Figure 6)

In this example, the pAN52-3 plasmid (described in Punt, P. J. et al., Journal of Biotechnology, 1990, vol. 17, pp. 19-34; called "starting plasmid" hereinafter) was the starting plasmid for construction of the expression vector in filamentous fungi (pThII) used to transform Penicillium roquefortii. Ligating the synthetic gene to this starting plasmid was performed in three stages described below.

(1.2.1) Ligating the 3/4 fragment

Thirty micrograms of pTZ18RN(3/4) was cut sequentially with Nco I and Hind III, generating 2 fragments. The small fragment with 310 bp containing the second part of the synthetic gene was purified in a 2% agarose gel. At the same time, 5 μg of the starting plasmid was cut sequentially with Nco I and Hind III. It was then dephosphorylated with alkaline phosphatase and a 5.8 Kb fragment was isolated in a 0.8% agarose gel. Then the starting plasmid, cut with Nco I and Hind III, dephosphorylated and purified, was ligated with the fragment of 310 bp from pTZ18RN(3/4). The mixture was used to transform E. coli DH5αF' as outlined in Figure 6. The clones containing the desired construction were identified by cutting the recombinant plasmids pTh(3/4) with Nco I and Hind III.

(1.2.2) Ligating fragment 1/2

In a second stage, plasmid pTZ18RN(1/2) was cut with Nco I and a NcoI-NcoI fragment containing the first part of the gene was purified in a 4% agarose gel. Plasmid pTh(3/4) was linearized with Nco I and processed with alkaline phosphatase. It was then ligated with the NcoI-NcoI fragment from pTZ18RN(1/2). The resulting plasmid was called pTh.

To analyze the clones, the pTh plasmid was cut with Bal I and Hind III. In the clones with the appropriate orientation, a fragment of 625 bp was obtained while those with inappropriate orientation produced a fragment of 300 bp.

(1.2.3) Ligating with the fungal marker

The pTh plasmid was then cut with Eco RI and the 5' ends were filled with the Klenow fragment of DNA polymerase I. This treated plasmid was then purified in a 0.8% agarose gel.

Starting with plasmid pEcoliR388 (N Datta, Saint Mary's Hospital, London), the sequence of resistance to sulfanilamide was obtained and a construction was made eliminating the procaryote promotor and terminator; then the structural gene was placed under the control of a promotor and a terminator of filamentous fungi (TrpC). The sulfanilamide resistance sequence obtained in this way was cut with SmaI and XbaI; the 5' ends were filled with Klenow and dNTP and a 1.75 Kb fragment was isolated from a 4% agarose gel. Then the fragment obtained in this way was ligated with pTh and transformation was carried out in E. coli DH1. The resulting plasmid was called pThII. This plasmid contains: (i) the synthetic gene which codifies thaumatin II under the control of a fungal promotor, and (ii) a sulfanilamide resistance marker. The final identity of the plasmid was verified by sequentiation as described in section 1.3.3.

(1.3) Transformation of Penicillium roquefortii with the aforementioned fungal expression vector

(1.3.1) Protoplast preparation

The protoplasts of Penicillium roquefortii used in the transformation experiments were prepared according to the following process, starting with the MUCL 29148 strain. Its conidia were inoculated in 50 mL of MSDPM liquid medium (medium semi-defined for mycelium production, the composition of which is described below). The culture was incubated for 44 hours at 28 °C in a mechanical stirrer at 270 rpm. The mycelium was recovered by filtration, washed with sterile water and resuspended in a 1.2M KCl solution containing 40 mg of Lysin Enzyme (Sigma) per gram of mycelium. After 4 hours of incubation at 28 °C at moderate stirring speed, protoplasts were obtained. Cell debris was eliminated by glass wool filtration. The protoplast suspension was washed and centrifuged (2000 rpm, 10 min.) twice with a 1.2 M KCl solution (10 mL/g). Finally, the protoplasts were resuspended in 1.2 M KCl (1 mL/g). This protoplast suspension (10^7-10^8 prot/mL) was used for the transformation experiments.

(1.3.2) Transformation

The protoplasts were centrifuged (2000 rpm, 10 min.) and then resuspended (5 x 10^8 protoplasts/mL) in solution I: 1.2 M KCl; 50 mM Tris.HCl (pH 8), 50 mM $CaCl_2$ and 20% of solution II (see below). They were incubated for 10 minutes at 28 °C. Aliquots of 0.1 mL were mixed with DNA (10 μg) from the expression plasmid, which contained the thaumatin II gene. Immediately afterward, 2 mL of solution II [1.2 M KCl; 50 mM Tris.HCl (pH 8), 50 mM $CaCl_2$ and 30% PEG 6000] was added. This mixture was incubated for 5 minutes at room temperature. After recovering the protoplasts by centrifugation (2000 rpm, 10 min.), they

were resuspended in 1 mL of 1.2 M KCl. Finally, aliquots of the protoplasts treated in this way were replated in petri dishes containing an appropriate medium for regeneration of the cell wall and subsequent selection using sulfanilamide (750 µg/mL). Using this transformation method, various strains that are resistant to sulfanilamide were isolated. These strains were analyzed to verify if the synthetic gene of thaumatin II had been incorporated into its genome.

(1.4) Analysis of the transformants

(1.4.1) PCR analysis

Analysis of the transformants obtained as described above to detect the DNA sequences of the synthetic gene of thaumatin II and resistance to sulfanilamide was performed using standard PCR techniques with appropriate oligonucleotides. Specifically, the T1 and T2 oligonucleotides were used, the sequences of which are included in section (1.4.1.2). T1 is complementary to nucleotides 605 and 624 of the upper strand of the synthetic gene of thaumatin II, while T2 is complementary to nucleotides 21 to 46 of the lower strand. Therefore, with these two oligonucleotides it was possible to amplify a fragment of 604 pairs of bases corresponding to oligonucleotides 21 to 624 of the synthetic gene of thaumatin II.

Figure 7 shows the success of the results, indicating that in the untransformed fungus (control), no bands appear of the size corresponding to the synthetic gene (lane 2), while in two of the transformant genes (M0901 and T0901) bands appear with the same number of bases as the band corresponding to the synthetic gene inserted in the pThII plasmid (lanes 3 to 5).

(1.4.1.1) Extraction of nucleic acids

The starting material was 5 g of mycelium which had been vacuum filtered using a Büchner funnel and which came from a 5-day MSDPM culture (0.6% $NaNo_3$; 0.052% $MgSO_4 \cdot 7H_2O$; 0.052 KCl; 1% glucose; 0.5% yeast extract; 0.5% casamino acids; $FeSO_4 \cdot 7H_2O$ traces; $ZnSO_4 \cdot 7H_2O$ traces).

The mycelium was ground in liquid nitrogen with a porcelain mortar. The mycelium was resuspended in the extraction buffer (10 mM Hepes, pH 6.9; 0.3 M saccharose; 20 mM EDTA, pH 8.0; 0.5% SDS) at a ratio of 10 mL of buffer per gram of mycelium. It was incubated for 15 minutes at 65 °C and centrifuged for 5 minutes at 7000 rpm (Beckman JA20 rotor) at room temperature to eliminate cell debris; the supernatant was collected and treated twice with phenol/chloroform/isoamyl alcohol (49:49:2) to eliminate proteins. The aqueous phase was precipitated with 0.3 M sodium acetate and 2.5 volumes of ethanol for 20 minutes at -20 °C. The precipitated volume was centrifuged at 7000 rpm for 20 minutes. The precipitate was resuspended in 1 mL of TE buffer, pH 8.0.

(1.4.1.2) PCR reaction mix

In a total volume of 100 µL, 20 ng of DNA and 10 µL of PEC 10X buffer were mixed (500 mM KCl; 15 mM $MgCl_2$; 100 mM Tris HCl, pH 8.3; 0.01% porcine gelatin; a mixture of DNTPs, with a concentration of 200 µM of each; 2.5 units of Amplitaq and 1 µM of primer). The synthetic oligonucleotides used were T1 (26 nucleotides) and T2 (20 nucleotides) and specific primers for the beginning and end of the synthetic gene of thaumatin II.

T1: 5'-CCGCTGCTCCTACACCGTCTGGGCCG-3'

T2: 5'-TTAGGCGGTGGGGCAGAAGG-3'

Twenty µL of mineral oil was added to the mixture to keep the sample from evaporating.

(1.4.1.3) PCR

The sample underwent a cycle at 94 °C for 5 minutes to separate the two DNA strands. Thirty chain reactions were then performed: first the DNA was denatured for 1 minute at 94 °C; the temperature was lowered to 55 °C for 30 seconds to allow the specific primers to join with the denatured DNA strand; the temperature was then increased again to 72 °C for 1 minute to allow the new strand (in formation) to elongate. When all the cycles were completed, a final elongation was performed for 5 minutes at 72 °C. The products of each PCR were analyzed in 0.8% agarose gel (Figure 7). Using this method two strains were identified called M0901 and T0901, the genomes of which contained the synthetic gene of thaumatin II.

(1.4.2) Immunoblotting Detection (Western-Blot)

Once the transformants that had incorporated themselves into the thaumatin II gene were detected correctly, Western blot was performed on the expression (Burnette W.N., Analytical Biochemistry, 1981, vol. 112, pp. 195-203), using polyclonal antibodies which had been previously obtained through standard rabbit immunization techniques to identify the protein. The serum obtained from each rabbit was precipitated with ammonium sulphate using standard techniques to precipitate the immunoglobulins, thus producing a protein fraction enriched with IgG antibodies. Figure 8 shows the outcome of the results obtained, indicating that no bands of the size corresponding to thaumatin II appear in the untransformed fungus (control), while in two of the transformed fungi a band appears having the same molecular weight as commercial thaumatin II.

(1.4.2.1) Preparation of the samples

The starting material was 2 g of mycelium which had been vacuum filtered using a Büchner funnel and which came from a 5-day culture at 28 °C in MSDPM medium. Both the mycelium retained in the funnel (solid fraction) and in the culture medium (liquid fraction) were analyzed.

Solid Fraction

Ten mL of sonication solution (625 mM Tris.HCl, pH 6.5, 1mM PMSF, 5% β-mercaptoethanol) per gram of mycelium was added to the mycelium retained in the funnel. The mycelium was sonicated for 1 minute with 1-second pulses (i.e., 1 second sonificated, 1 second without sonification, and so on). The process was repeated three more times at intervals of from 3 to 5 minutes. It was centrifuged at 7500 rpm (Beckman JA20 rotor) for 20 minutes at 4 °C.

Liquid Fraction

β-Mercaptoethanol (final concentration 5%) and PMSF (final concentration 1 mM) were added to 3 mL of the extracellular medium. Three mL of both fractions was used to start and was concentrated by column centrifugation (Bio-Rad ultrafilters) which retain the proteins having a molecular weight greater than 10,000 Daltons. In this process, the 3 mL passing through the columns was reduced to 200 µL.

Twenty µL of the 2 x sample buffer (25% glycerol; 2.5% SDS; 0.25M Tris.HCl, pH 7.0; 10 mM EDTA, pH 8.0; 0.002% bromophenol blue) was added to 20 µL of the concentrated solutions. They were boiled for 5 minutes and immediately placed in protein denaturing gel (SDS-polyacrilamide).

The protein gels used were 14% polyacrilamide and 18% urea. Electrophoresis was performed at 150 volts and stopped when the front of the sample was 3 or 5 mm from the end of the gel.

(1.4.2.2) Transfer to nitrocellulose

Once the eletrophoresis was completed and after removing the piled-up part, the gel was transferred to nitrocellulose paper (NC). To do so, the Bio-Rad Trans-blot SD Semidry Unit was used. Transfer took 30 minutes at 15 volts.

Once the bands were transferred to NC paper, the paper was left in blocking solution (3% BSA; 0.01% sodium aside; 0.05% Tween-20 in TBS; TBS = 150 mM NaCl; 50 mM Tris.HCl, pH 8.0) and stirred overnight. After this operation, the NC paper was processed as follows.

The NC paper was taken out of the blocking solution, washed with TBS and incubated with serum: immune IgG fraction (0.37 mg/mL) diluted (1:500) in blocking solution (with sodium azide). As a negative control, the normal preimmune IgG fraction was used (0.35 mg/mL) diluted (1:500) in blocking solution (with sodium azide). The solution was stirred and incubated for 4 hours at room temperature.

Three 10-minute washes were performed in TBS-Tween (TBS 1X + Tween-20, 0.05%). It was stirred and incubated for 4 hours at room temperature with the secondary antibody: anti-rabbit IgG-phosphatase alkaline conjugate diluted (1:500) in blocking solution (without sodium azide). Three 10-minute washes were performed in TBS-Tween.

The alkaline phosphatase reaction was performed: a) the NC was equilibrated with alkaline phosphatase buffer (100 mM Tris.HCl, pH 9.5 100; 100mM NaCl; 50 mM $MgCl_2$); b) the NC was placed in the development reaction mix (15 mL of alkaline phosphatase buffer, 66 µL of nitro blue tetrazodium, NBT) (75 mg/mL in 70% dimethyl formamide), 99 µL of 5-bromo-4-chloro-3-indole phosphate (BCIP) (25 mg/mL in 100% dimethyl formamide) until the bands turned dark; c) the reaction was stopped with alkaline phosphate

stop solution (20 mM Tris.HCl, pH 8.0 and 20 mM EDTa, pH 8.0).

(1.4.2.3) Protein gel staining

The gels were stained for 1 hour with staining solution and stirred gently (25 ethanol; 10% acetic acid; 0.1% Comassie blue). They were destained with destaining solution (25% methanol; 7.5% acetic acid) until the blue color faded from the gel base.

EXAMPLE 2: EXTRACELLULAR PRODUCTION OF THAUMATIN IN PENICILLIUM ROQUEFORTII

For extracellular production of thaumatin, Penicillium roquefortii was transformed with plasmid pThIII, which was constructed as described below and outlined in Figure 9.

Plasmid pThII, described above, (section 1.2.3) was purified using standard techniques and resuspended in TE buffer at a final concentration of 1 µg/µl. Thirty micrograms (µg) of this plasmid were cut with restriction enzymes MscI and HindIII, and a fragment of 646 base pairs containing the complete gene of thaumatin II was isolated in a 0.8% agarose gel. The ends of the fragment were converted to blunt ends with the Klenow fragment from DNA polymerase I.

Plasmid pAN52-6B, containing approximately 7.5 Kb and derived from pAN52-6 Not 1 (cf. Van den Hondel et al., "Heterologous Gene Expression in filamentous fungi"; in Bennett and Lasvre, "More Gene Manipulation in Fungi"; Academic Press, 1991, chapter 18, pp. 396-428) was digested with BssHII and its ends were converted to blunt ends through the action of the Klenow fragment of DNA polymerase I.

These two fragments were ligated using DNA ligase and the resulting mix was used to transform the DH5αF' strain of E. coli. The resulting plasmid, pThII-bis, was isolated and its structure verified by sequencing using the Sanger dideoxy method.

The following step was to cut the pThII-bis plasmid (8.1 Kb) with XbaI and to isolate a fragment of approximately 5.5 Kb in length containing the thaumatin gene and the promoter sequence and glucoamylase signal sequence of Aspergillus niger. The trpC terminator sequence of Aspergillus nidulans was also present in this fragment.

The aforementioned 5.5 Kb fragment was ligated with a plasmid containing the sulfanilamide resistance sequence, previously cut with XbaI (the only cutting site on this plasmid). The ligating mix was used to transform E. coli. strain DH5αF'. The resulting plasmid was called pThIII, as indicated in Figure 9.

The pThIII plasmid contained: (i) the synthetic gene which codifies thaumatin II under the control of the glucoamylase promoter of Aspergillus niger; (ii) the signal sequence ("pre") and the "pro" sequence of the glucoamylase gene of Aspergillus niger; (iii) a sulfanilamide resistance marker; and (iv) the trpC terminator of Aspergillus nidulans. The final identity of this construction was verified by sequencing.

A strain of Penicillium roquefortii was transformed with plasmid pThIII according to the same method described in Example 1 (sections 1.3.1 and 1.3.2). The colonies resistant to sulfanilamide were tested to see if their genomes contained the substantially modified gene codifying thaumatin II. The methods used (PCR) were analogous to those described in Example 1 (section 1.4.1).

Figure 10 shows the result of a PCR experiment. The two oligonucleotides used to detect the thaumatin gene were the same ones used before (T1 and T2). With these two oligonucleotides, a fragment of 604 pairs of bases can be amplified corresponding to nucleotides 21 to 624 of the synthetic gene encoding thaumatin II. Figure 10 shows that when DNA from an untransformed fungus ("control", lane 2) is used, none of the bands corresponding to the synthetic gene are amplified, whereas when DNA is used from a fungus transformed with pThIII, a band of the expected size is amplified (lane 3). This fungus was called transformant a2. For control purposes, the reaction products obtained when plasmid pThIII was used were also run through the gel (lane 4).

The figure shows that transformant a2 correctly incorporated the synthetic gene of thaumatin II in its genome. Therefore, it was analyzed in greater detail to see if it expressed and secreted thaumatin II correctly. For immunoblotting analysis (Western-Blot) of the recombinant thaumatin, the methods described in section (1.4.2.) were used with the following modifications.

The experiment was started with 1 liter of a2 strain of Penicillium roquefortii which was grown for 8 days at 28°C in a semidefined medium for mycelium production (MSDPM). After vacuum filtration with a Büchner funnel, producing 45 g of mycelium per liter of culture, both the culture medium (liquid fraction) and the retained mycelium (solid fraction, 4.5 g) were analyzed.

The solid fraction was processed using the methods outlined in section (1.4.2.1), including sonication, thus obtaining 13.5 mL of mycelium extract in sonication solution.

The 13.5 mL of mycelium extract and 10 mL of culture medium were precipitated with 10% trichloracetic acid and the precipitated material was resuspended in a final volume of 200 µL of sonication solution. These samples were then analyzed by protein electrophoreses and immunoblotting as described in detail in Example 1, section (1.4.2).

The results of this experiment are shown in Figure 11 (14% SDS-polyacrylamide gel). Lane 7 in this figure contains proteins of known molecular weight (markers). The molecular weight corresponding to each protein is listed on the right of the figure. Lane 2 contains a sample of culture medium where fungus T0901 was grown. As described in Example 1, this fungus is a producer of intracellular thaumatin. Lanes 3 and 5 contain samples of culture medium (E for extracellular) and mycelium (I for intracellular) corresponding to transformant a2. Lanes 4 and 6 contain the same samples (E and I) corresponding to untransformed Penicillium roquefortii. As is clearly seen in Figure 11, transformant a2 turned out to be a good producer and secretor of thaumatin.

However, the effectiveness of the secretion was not complete given that a part of the thaumatin produced was not secreted, as is seen in the comparison between lanes 3 and 5. Organoleptic tests were performed on the culture broth and the characteristic sweet taste of thaumatin was detected.

EXAMPLE 3: EXTRACELLULAR PRODUCTION OF THAUMATIN IN THE AWAMORI VARIANT OF ASPERGILLUS NIGER

Strain NRRL312 of the awamori variant of Aspergillus niger was transformed in the presence of polyethylene glycol, as described in the literature (Yelton et al., Proc. Natl. Acad. Sci. USA, 1984, vol. 81, pp. 1470-4), with some modifications.

Four hundred mL of CM medium (malt extract, 5 g/L; yeast extract, 5 g/L; glucose, 5 g/L) in a 2-liter flask was inoculated with spores of the awamori variant of Aspergillus niger from a dish. The fungus grew for 16 hours. The mycelium was collected by filtration through a sterile gauze and washed with 100 mL of wash buffer (0.6 M $MgSO_4$, 10 mM Na_3PO_4, pH 5.8). The mycelium was pressed in sterile paper towels and produced 2.5 grams.

For the formation of protoplasts, the mycelium was resuspended in 15 mL/g of cold protoplast buffer (1.2 M $MgSO_4$, 10 mM Na_3PO_4, pH 5.8). At this point, 40 mg of Lysin enzyme (Sigma) was added per g of mycelium and the mixture was placed in ice for five minutes. After this incubation, 1 mL of BSA solution was added (12 mg/mL in protoplast buffer) and the solution was incubated for 3 or 4 hours at 30 °C. Protoplast formation was monitored using a microscope. The mixture was filtered through nylon or a glass membrane and washed with the protoplast buffer. The protoplasts were centrifuged at 2000 rpm at 4 °C for 15 minutes with a floating rotor (Beckman GPR centrifuge). The protoplasts were resuspended in 15 mL of ST solution (1M sorbitol, 10 mN Tris-HCl, pH 7.5), centrifuged again and resuspended in 1 mL of ST. The solution was centrifuged again and washed twice with 1 mL of STC (ST plus 0.01 M $CaCl_2$). The protoplasts were counted under the microscope, centrifuged again and resuspended in sufficient volume of STC to obtain a concentration of 10^8 protoplasts/mL. Each 400-mL culture generally produced 10^8 protoplasts. At that point, the protoplasts were directly plated in regeneration medium, in 5-mL tubes of 0.7% soft agar with saccharose osmotic stabilizer (1M), and were plated in basal medium with 1.5% agar.

For the transformation experiments, 200 µL of the 10^8-protoplasts/mL protoplast solution was used to start. Ten µg of transformant DNA (pThIII in this case) and 50 µL of PTC (60% PEG 6000; 10 mN Tris-HCl, pH 7.5; 10 mM $CaCl_2$) were added to the protoplasts and the solution was incubated in ice for 20 minutes. One mL of PTC was then added and the solution was mixed well and kept at room temperature for five minutes. The protoplasts were centrifuged and resuspended in 200 µL of STC medium. The mixture was plated in regeneration medium with sulfanilamide at 1 mg/mL. The dishes were incubated upside down at 30 °C. Regeneration was observed after three or four days of incubation.

(3.1) Preparation of the regeneration medium

1. Trace solution: 400 mg/L $CuSO_4 \cdot 5H_2O$; 800 mg/L $FeSO_4 \cdot 7H_2O$; 800 mg/L $MnSO_4 \cdot 2H_2O$; 800 mg/L $Na_2MoO_4 \cdot 2H_2O$; 40 mg/L $Na_2BrO_7 \cdot 10H_2O$; 8 mg/L $ZnSO_4 \cdot 7H_2O$.
2. Salt solution (50X): 26 g/L KCl; 26 g/L $MgSO_4 \cdot 7H_2O$; 76 g/L KH_2PO_4; 50 mL/L of trace solution.
3. Ammonium tartrate: 30 grams per liter.
4. MMA (minimum Aspergillus medium): 10 or 15 g of glucose, or 7 g of agar was added to 970 mL of distilled water (final concentrations of 1.5% or 0.7%, respectively). The mixture was autoclaved and 10 mL of sterile ammonium tartrate solution and 20 mL of sterile salt solution were then added. Finally, the regeneration medium was prepared by adding saccharose to the MMA medium until the concentration of

1 M was reached.

EXAMPLE IV: PRODUCTION, SECRETION AND PROCESSING OF A GLUCOAMYLASE-THAUMATIN FUSION PROTEIN

As outlined in Figure 12, the pGEX-KG plasmid (5.0 Kb) (Pharmacia Biotech) was sequentially controlled with NcoI and Hind III, thus generating a fragment of approximately 4900 bp. This fragment, which no longer contained the SalI restriction site of the pGEX-KG polylinker, was purified in a 0.8% agarose gel.

The previous fragment was ligated with a NcoI-HindIII fragment from plasmid pTZ18RN(3/4) which contained the second part of the synthetic gene of thaumatin, thus generating plasmid pECThI of approximately 5.3 Kb. This new plasmid was treated with NcoI and the linearized fragment was ligated with a NcoI-NcoI fragment from plasmid pTZ18RN(1/2), which contained the first part of the synthetic gene of thaumatin, thus generating plasmid pECThII (of approximately 5.6 Kb). Plasmid pECThII contained the synthetic gene of thaumatin under the control of the tac promoter of Escherichia coli. This construction made it possible to obtain intracellular production of recombinant thaumatin in Escherichia coli.

The starting point for the construction of pThIX was the pECThI plasmid (of approximately 5.3 Kb). To eliminate the only MscI restriction site present in this plasmid, pECThI was sequentially treated with MscI and EcoRV (enzymes which produce blunt ends), thus releasing two fragments of 4.1 Kb and 1.2 Kb. The 4.1-Kb fragment was purified in a 0.8% agarose gel and religated through the action of DNA ligase. The result was plasmid pThIV. This plasmid was linearized with NcoI and the linear fragment was ligated with a NcoI-NcoI fragment from plasmid pTZ18RN(1/2), which contained the first half of the synthetic gene of thaumatin, thus generating plasmid pThV.

The single-stranded oligonucleotides, GLA1 and GLA2, were commercially bought (Ingenasa S.A) and have the following sequences (included in those of Figure 14):

GLA1: 5'-AATTCTGCGGAACGTCGACCGCGACGGTGACTGACACCTGGCGGC GAATGGATAAAAGGG-3'

GLA2: 5'-CCCTTTTATCCATTCGCCGCCAGGTGTCAGTCACCGTCGCGGTCG ACGTTCCGCAG-3'

These two oligonucleotides were annealed as follows: 10µg of each oligonucleotide was mixed in ligation buffer (40 mM Tris-HCl, pH 7.5; 20 mM $MgCl_2$; 50 mM NaCl) in a final volume of 25 µL. The mixture was heated for 5 minutes at 65 °C and the temperature was allowed to drop slowly (for one half hour) to 30 °C. The double-stranded DNA annealed in this way was purified in a 8% polyacrilamide gel. This double-stranded synthetic oligonucleotide, called GLA(1/2), had one blunt edge and one EcoRI end. Plasmid pThV was digested with MscI and EcoRI and ligated with the GLA(1/2) synthetic fragment, thus generating pThVI. Figure 14 shows the connection between the last sequences of the glucoamylase gene of Aspergillus niger, the spacer sequence and the synthetic gene of thaumatin II.

The next step was to insert the complete gene of glucoamylase (glaA) of Aspergillus niger or the awamori variant of Aspergillus niger, respectively, in phase with the complete gene of thaumatin II so that a glucoamylase-thaumatin fusion protein could be formed.

Plasmid pFGA2, obtained from the Belgian collection of cultures and LMBP plasmids (Ghent, Belgium, number 1728), contained the complete gene of glucoamylase (glaA) of Aspergillus niger. The plasmid was digested with EcoRI and SalI, and a fragment of approximately 2.3 Kb was isolated containing the complete gene of glucoamylase except for the last 10 amino acids of the protein. This fragment was ligated with plasmid pThVI which had previously been digested with EcoRI and SalI, thus generating plasmid pThVII (the junctions are described in Figure 14).

To obtain the glucoamylase gene of the awamori variant of Aspergillus niger, the following process was followed: total DNA of the NRRL312 strain of this fungus was prepared according to the protocol in section (1.4.1.1) . Two oligonucleotides, complementary to the 5' and 3' ends of the glucoamylase gene were used to amplify the complete gene. The fragment thus amplified was purified in a 0.8% agarose gel and digested with EcoRI and SalI. This 2.3-Kb EcoRI-SalI fragment was subcloned in pBluescript SK (Stratagene Inc.), which had previously been digested with EcoRI and SalI, thus generating plasmid pGLA-Aw.

In order to place the glucoamylase-spacer-thaumatin cassette under the control of the gla promoter of Aspergillus niger, the pThVII plasmid was digested with the restriction enzymes BssHII (partial digestion) and HindIII, and a fragment of approximately 3.0 Kb was isolated. This fragment was ligated with pAN52-6B which had previously been digested with BssHII and HindIII, thus obtaining plasmid pThVIII. Finally, the sulfanilamide resistance gene (Su^R) was inserted as described in Example 2, thus generating pThIX.

Plasmid pThIX contained: (i) a sulfanilamide resistance marker; (ii) a DNA sequence which encodes a fusion protein formed by (a) the synthetic gene of thaumatin II, (b) a spacer sequence which in turn contains a KEX2 processing sequence, and (c) the complete glucoamylase gene of Aspergillus niger; and finally, (iii)

the signal sequence ("pre") and the "pro" sequence of the glucoamylase gene (glaA) of Aspergillus niger.

Plasmid pThIX was used to transform the awamori variant of Aspergillus niger as per the protocols specified in Example 3. Transformants which correctly secreted and processed thaumatin were obtained, and it was determined that the protein was sweet.

In the same way, but using plasmid pGLA-Aw instead of plasmid pThVII, an analogue plasmid of pThIX was obtained containing the gla sequence of A. awamori instead of that of A. niger. Similarly, this plasmid was also used to transform a strain of A. awamori, with similar results.

LIST OF SEQUENCES

SEQ. ID No.1

```
GCC ACC TTC GAG ATC GTC AAC CGC TGC TCC TAC ACC GTG TGG GCG GCC  48
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
1               5                  10                  15
GCC TCC AAA GGC GAC GCC GCC CTG GAC GCC GGC GGC CGC CAG CTC AAC  96
Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30
TCG GGA GAG TCC TGG ACC ATC AAC GTA GAA CCC GGC ACC AAG GGC GGC 144
Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Lys Gly Gly
        35                  40                  45
AAA ATC TGG GCC CGC ACC GAC TGC TAT TTC GAC GAC AGC GGC CGC GGC 192
Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Arg Gly
    50                  55                  60
ATC TGC CGG ACC GGC GAC TGC GGC GGC CTC CTC CAG TGC AAG CGC TTC 240
Ile Cys Rag Thr Gly Asp Cys Gly Gly Leu Leu Gln Cys Lys Arg Phe
65                  70                  75                  80
GGC CGG CCG CCC ACC ACG CTG GCG GAC TTC TCG CTC AAC CAG TAC GGC 288
Gly Arg Pro Pro Thr Thr Leu Ala Glu Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95
AAG GAC TAC ATC GAC ATC TCC AAC ATC AAA GGC TTC AAC GTG CCG ATG 336
Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110
GAC TTC TGC CCG ACC ACG CGC GGC TGC CGC GGG GTG CGG TGC GCC GCC 384
Asp Phe Ser Pro Thr Thr Arg Gly Cys Arg Gly Val Arg Cys Ala Ala
        115                 120                 125
GAC ATC GTG GGC CAG TGC CCG GCG AAG CTG AAG GCG CCG GGG GGT GGT 432
Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140
TGC AAC GAT GCG TGC ACC GTG TTC CAG ACG AGC GAG TAC TGC TGC ACC 480
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160
ACG GGG AAG TGC GGG CCG ACG GAG TAC TCG CGC TTC TTC AAG AGG CTT 528
Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175
TGC CCC GAC GCG TTC AGT TAT GTC CTG GAC AAG CCA ACC ACC GTC ACC 576
Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190
TGC CCC GGC AGC TCC AAC TAC AGC GTC ACT TTC TGC CCT ACT GCC(TAA)n 624
Cys Pro Gly Ser Ser Asn Tyr Arg Val Thr Phe Cys Pro Thr Ala
        195                 200                 205
```

Sequence ID No. 1: Amino-acid sequence of the protein thaumatin II, and nucleotide sequence of the natural gene.

SEQ. ID No.2

```
GCC ACC TTC GAG ATC GTC AAC CGC TGC TCC TAC ACC GTC TGG GCC GCC  48
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15
GCC TCC AAG GGC GAC GCC GCC CTC GAC GCC GGC GGC CGC CAG CTC AAC  96
Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30
TCC GGC GAG TCC TGG ACC ATC AAC GTC GAG CCC GGC ACC AAG GGC GGC 144
Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Lys Gly Gly
        35                  40                  45
AAG ATC TGG GCC CGC ACC GAC TGC TAC TTC GAC GAC TCC GGC CGC GGC 192
Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Arg Gly
    50                  55                  60
ATC TGC CGC ACC GGC GAC TGC GGC GGC CTC CTC CAG TGC AAG CGC TTC 240
Ile Cys Arg Thr Gly Asp Cys Gly Gly Leu Leu Gln Cys Lys Arg Phe
65                  70                  75                  80
GGC CGC CCC CCC ACC ACC CTC GCC GAC TTC TCC CTC AAC CAG TAC GGC 288
Gly Arg Pro Pro Thr Thr Leu Ala Glu Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95
AAG GAC TAC ATC GAC ATC TCC AAC ATC AAG GGC TTC AAC GTC CCC ATG 336
Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110
GAC TTC TCC CCC ACC ACC CGC GGC TGC CGC GGC GTC CGC TGC GCC GCC 384
Asp Phe Ser Pro Thr Thr Arg Gly Cys Arg Gly Val Arg Cys Ala Ala
        115                 120                 125
GAC ATC GTC GGC CAG TGC CCC GCC AAG CTC AAG GCC CCC GGC GGC GGC 432
Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140
TGC AAC GAC GCC TGC ACC GTC TTC CAG ACC TCC GAG TAC TGC TGC ACC 480
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160
ACC GGC AAG TGC GGC CCC ACC GAG TAC TCC CGC TTC TTC AAG CGC CTC 528
Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175
TGC CCC GAC GCC TTC TCC TAC GTC CTC GAC AAG CCC ACC ACC GTC ACC 576
Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190
TGC CCC GGC TCC TCC AAC TAC CGC GTC ACC TTC TGC CCC ACC GCC(TAA)n 624
Cys Pro Gly Ser Ser Asn Tyr Arg Val Thr Phe Cys Pro Thr Ala
        195                 200                 205
```

Sequence ID No. 2: Amino-acid sequence of thaumatin II and nucleotide sequence of the artificial, synthetic and completely optimized gene, used in the examples of this invention, to which the n codons with TAA termination ($n \geq 1$) were added.

Claims

1. A DNA sequence which codifies the amino-acid sequence corresponding to the 207 amino acids of the protein thaumatin II (included in Sequence ID No. 1), followed by n stop sequences, where integer n is greater than or equal to 1; said DNA sequence being characterized in that it is the result of modifying, more than 50% of the possible, the DNA sequence of the natural gene which codifies the 207 amino acids of thaumatin II (natural gene also shown in Sequence ID No. 1); said modification consisting of adding n stop codons, where integer n is greater than or equal to one, and effecting more than 50% of the possible changes in the nucleotide codons corresponding to the amino acids of thaumatin II; said changes consisting of replacing the original codons in all the amino acids possible, with the codons indicated in parentheses in the following list of amino-acid codons:
   Ala (GCC), Arg (CGC), Asn (AAC), Asp (GAC), Cys (TGC), Lys (AAG), Gln (CAG), Glu (GAG), Gly (GGC), Ile (ATC), Leu (CTC), Met (ATG), Phe (TTC), Pro (CCC), Ser (TCC), Thr (ACC), Trp (TGG), Tyr (TAC), Val (GTC).

2. A DNA sequence according to claim 1 where the modification consists of adding from one to three stop codons (n = 1, 2 or 3), and effecting more than 75% of the possible codon changes.

3. A DNA sequence according to claim 2 where all (100%) of the possible codon changes have been made so that the DNA sequence is the one in Sequence ID No. 2.

4. A DNA sequence according to any of claims 1, 2 or 3, wherein the stop codon(s) represent TAA.

5. A recombinant plasmid comprising: (i) a DNA sequence according to any of the claims 1 to 4; (ii) an expression cassette for filamentous fungi containing one promoter sequence and one terminating sequence for this type of fungi; (iii) an appropriate selection marker; and, optionally, (iv) a secretion signal DNA sequence for the extracellular production of the protein.

6. A recombinant plasmid according to claim 5 where the promoter sequence of the expression cassette comes from the gene of the enzyme glyceraldehyde 3-phosphate dehydrogenase of Aspergillus niger, or from the glucoamylase gene of the same fungus; the terminating sequence of the expression cassette is that of tryptophan C of Aspergillus nidulans; and the selection marker is the sulfanilamide resistance marker.

7. A recombinant plasmid expressing the fusion protein thaumatin-glucoamylase comprising: (i) an appropriate selection marker; (ii) a DNA sequence made up of (a) a DNA sequence according to any of the claims 1 to 4, (b) a spacer sequence which in turn contains a KEX2 processing sequence, and (c) the complete glucoamylase gene of Aspergillus niger or the awamori variant of Aspergillus niger (glaA); and (iii) the "pre" signal sequence and the "pro" sequence of the glaA gene.

8. A filamentous fungus culture capable of producing the protein thaumatin II, which has been transformed with any of the plasmids in claims 5 to 7.

9. A culture according to claim 8 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

10. A process for producing thaumatin II comprising the following steps:
a) insertion of the DNA sequence from claims 1, 2, 3 or 4 in any of the expression vectors in claims 5, 6 and 7, using standard recombinant DNA technology techniques;
b) transformation of a strain of filamentous fungus with this expression vector;
c) culture of the strain of filamentous fungus which has been transformed in this way under the appropriate nutrient conditions, thus producing thaumatin II, either intracellularly, extracellularly or in both ways simultaneously, or in the form of the fusion protein thaumatin-glucoamylase.
d) depending on the case, separation and purification of thaumatin II alone, or separation of thaumatin II from the culture medium together with the fungus mycelium.

11. A process according to claim 10 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

12. A DNA sequence which codifies the amino-acid sequence corresponding to the 207 amino acids of the protein thaumatin I (207 amino acids which differ from those of Sequence ID No. 1 in only five amino acids, namely, 46-Asn, 63-Ser, 67-Lys, 76-Arg and 113-Asn), characterized in that it has the following five fixed codons: AAC (46-Asn), TCC (63-Ser), AAG (67-Lys), CGC (76-Arg) and AAC (113-Asn), and the rest of the codons are as in the DNA sequence in claim 1.

13. A DNA sequence according to claim 12 which has from one to three stop codons (n = 1, 2 or 3), and the rest of the codons which differ from the five fixed ones, as in the DNA sequence in claim 2.

14. A DNA sequence according to claim 13 which has the codons which are different from the five fixed ones, as in the DNA sequence in claim 3.

15. A recombinant plasmid comprising: (i) a DNA sequence according to any of the claims 12, 13 or 14; (ii) an expression cassette for filamentous fungi containing a promoter sequence and a terminating sequence which are appropriate for this type of fungi; (iii) an appropriate selection marker; and, optionally, (iv) a secretion signal DNA sequence for the extracelluar production of the protein.

16. A recombinant plasmid according to claim 15 where the promoter sequence of the expression cassette comes from the gene of the enzyme glyceraldehyde 3-phosphate dehydrogenase of Aspergillus niger, or from the glucoamylase gene of the same fungus; the terminating sequence of the expression cassette is tryptophane C of Aspergillus nidulans; and the selection marker is the sulfanilamide resistance selection marker.

17. A recombinant plasmid which expresses the fusion protein thaumatin-glucoamylase comprising: (i) an appropriate selection marker; (ii) a DNA sequence made up of (a) a DNA sequence according to claims 12, 13 or 14, (b) a spacer sequence which in turn contains a KEX2 processing sequence, and (c) the complete gene of glucoamylase of Aspergillus niger or the awamori variant of Aspergillus niger (glaA); and (iii) the "pre" signal sequence and the "pro" sequence from the glaA gene.

18. A filamentous fungus culture capable of producing the protein thaumatin I, which has been transformed with any of the plasmids in claims 15 to 17.

19. A culture according to claim 18 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

20. A process for producing thaumatin I comprising the following steps:
a) insertion of the DNA sequence of any of the claims 12, 13 or 14 in one expression vector selected from those in claims 15, 16 and 17, using standard recombinant DNA technology techniques;
b) transformation of a strain of filamentous fungus with this expression vector;
c) culture of the strain of filamentous fungus which has been transformed in this way under the appropriate nutrient conditions, thus producing thaumatin I either intracellularly, extracellularly or in both ways simultaneously, or in the form of the thaumatin-glucoamylase fusion protein.
d) depending on the case, separation and purification of thaumatin I alone, or separation of thaumatin I from the culture medium together with the fungus mycelium.

21. A process according to claim 20 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

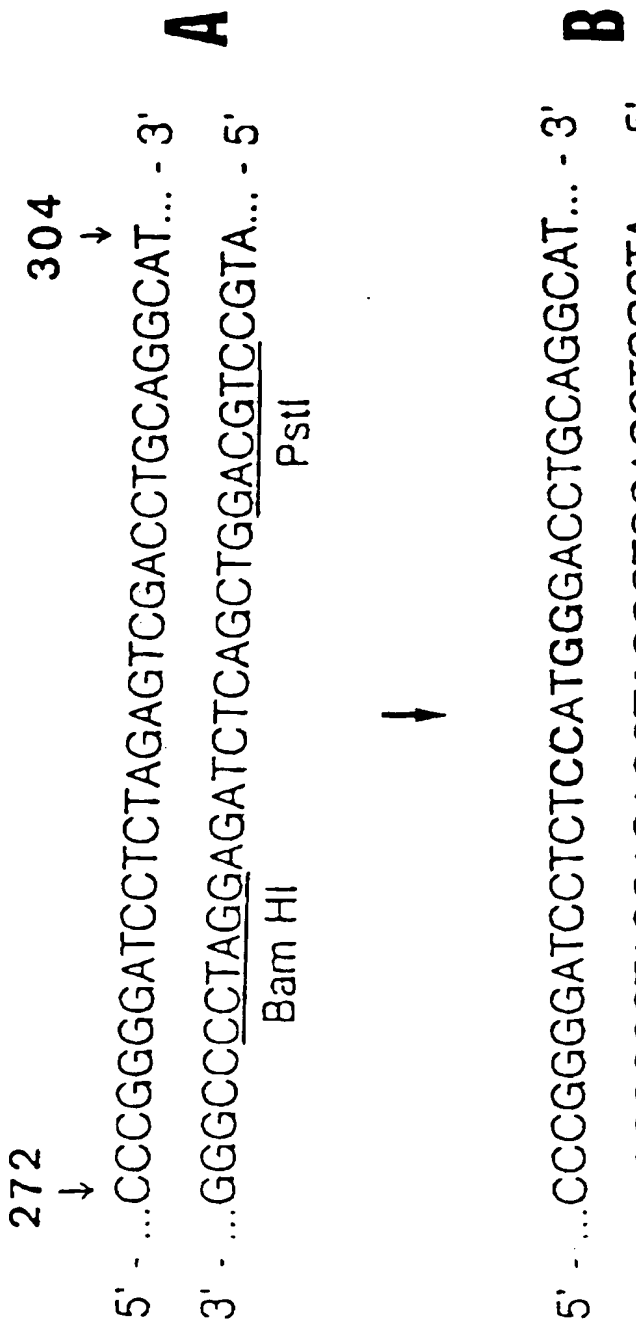


Fig. 1

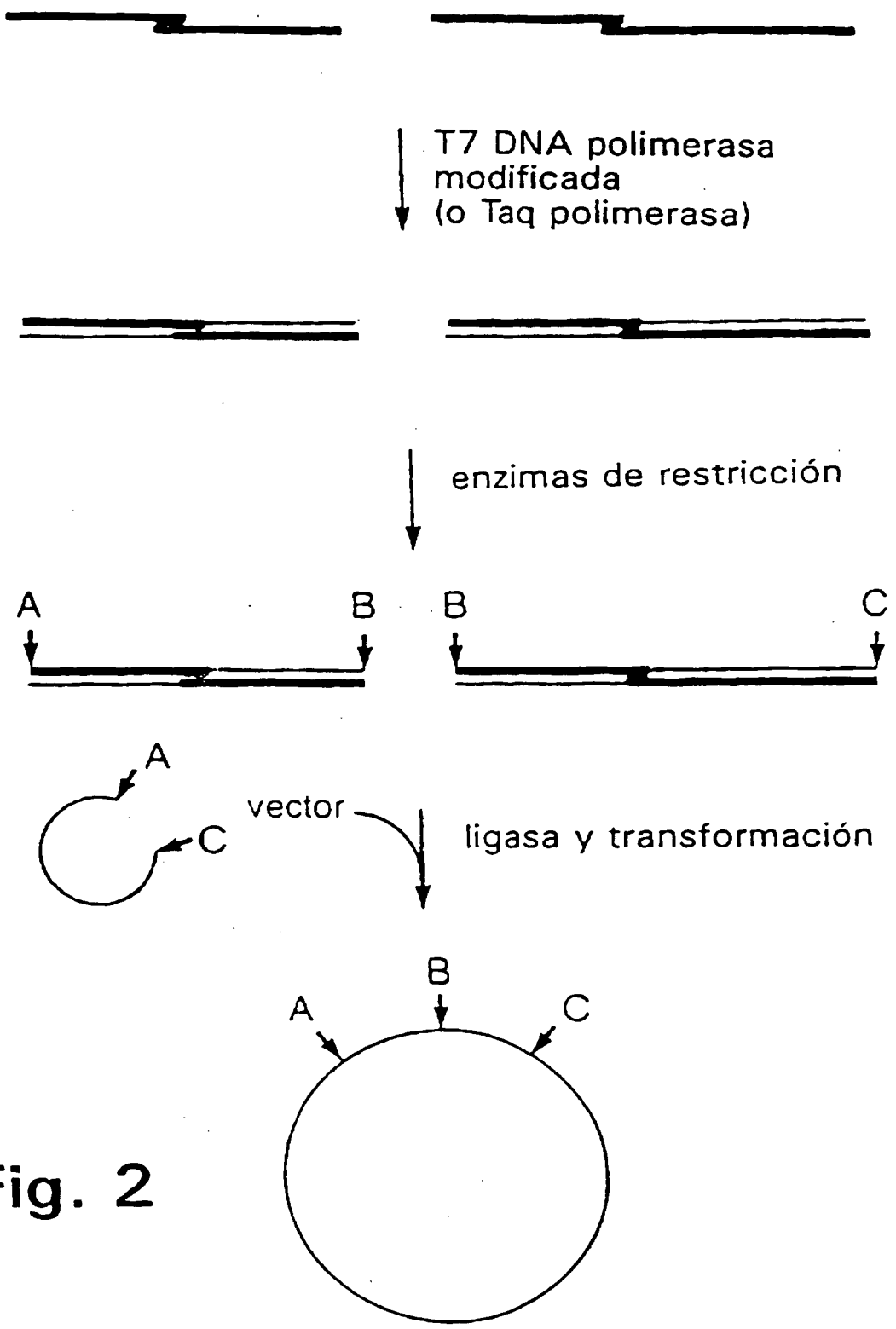


Fig. 2

1a (106 nucleótidos).

5'- AAATGGAGGATCCATGGCCACCTTCGAGATCGTCAACCGC
TGCTCCTACACCGTCTGGGCCGCCGCCTCCAAGGGCGACGC
CGCCCTCGACGCCGGCGGCCGCCAG - 3'

1b (87 nucleótidos).

5'- GCGGGCCCAGATCTTGCCGCCCTTGGTGCCGGGCTCGAC
GTTGATGGTCCAGGACTCGCCGGAGTTGAGCTGGCGGCCGC
CGGCGTC - 3'

2a (117 nucleótidos).

5'- GGCGGCAAGATCTGGGCCCGCACCGACTGCTACTTCGACG
ACTCCGGCCGCGGCATCTGCCGCACCGGCGACTGCGGCGGC
CTCCTCCAGTGCAAGCGCTTCGGCCGCCCCCCCACC - 3'

2b (103 nucleótidos).

5'- AGTCCATGGGGACGTTGAAGCCCTTGATGTTGGAGATGTCG
ATGTAGTCCTTGCCGTACTGGTTGAGGGAGAACTCGGCGAGGG
TGGTGGGGGGGCGGCCGAA - 3'

Fig. 3

3a (84 nucleótidos).

5'- AACGTCCCCATGGACTTCTCCCCCACCACCCGCGGCTGCCGCGGC
GTCCGCTGCGCCGCCGACATCGTCGGCCAGTGCCCCGCC - 3'

3b (64 nucleótidos).

5'- AGACGGTGCAGGCGTCGTTGCAGCCGCCGCCGGGGGCCTTGAGCT
TGGCGGGGCACTGGCCGAC - 3'

4a (101 nucleótidos).

5'- TCCAGACCTCCGAGTACTGCTGCACCACCGGCAAGTGCGGCCCCA
CCGAGTACTCCCGCTTCTTCAAGCGCCTCTGCCCCGACGCCTTCTCCT
ACGTCCTC - 3'

4b (107 nucleótidos).

5'- GCTTGCCTGCAGTTATTATTAGGCGGTGGGGCAGAAGGTGACGCGG
TAGTTGGAGGAGCCGGGGCAGGTGACGGTGGTGGGCTTGTCGAGGAC
GTAGGAGAAGGCGT - 3'

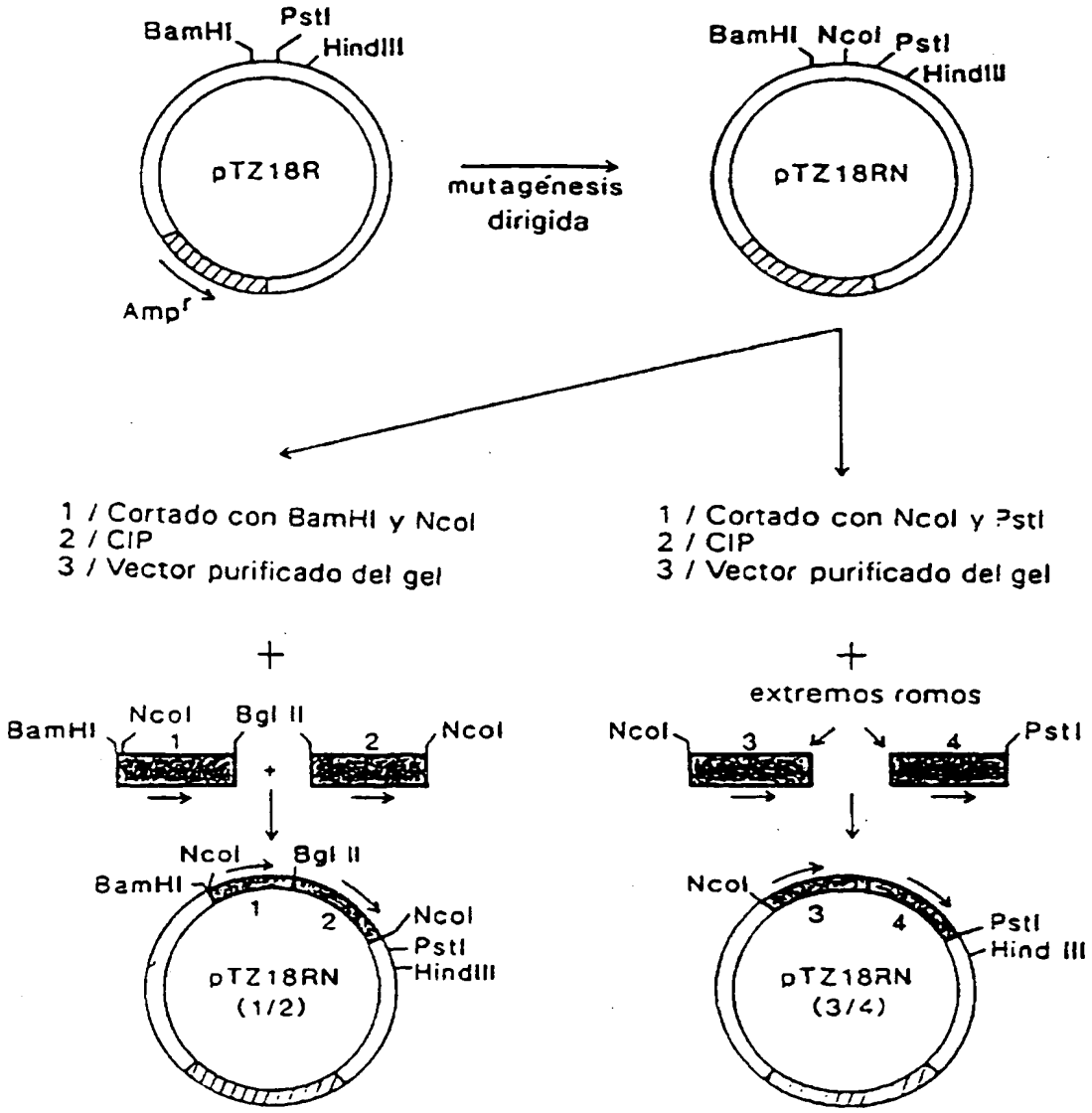


Fig. 4

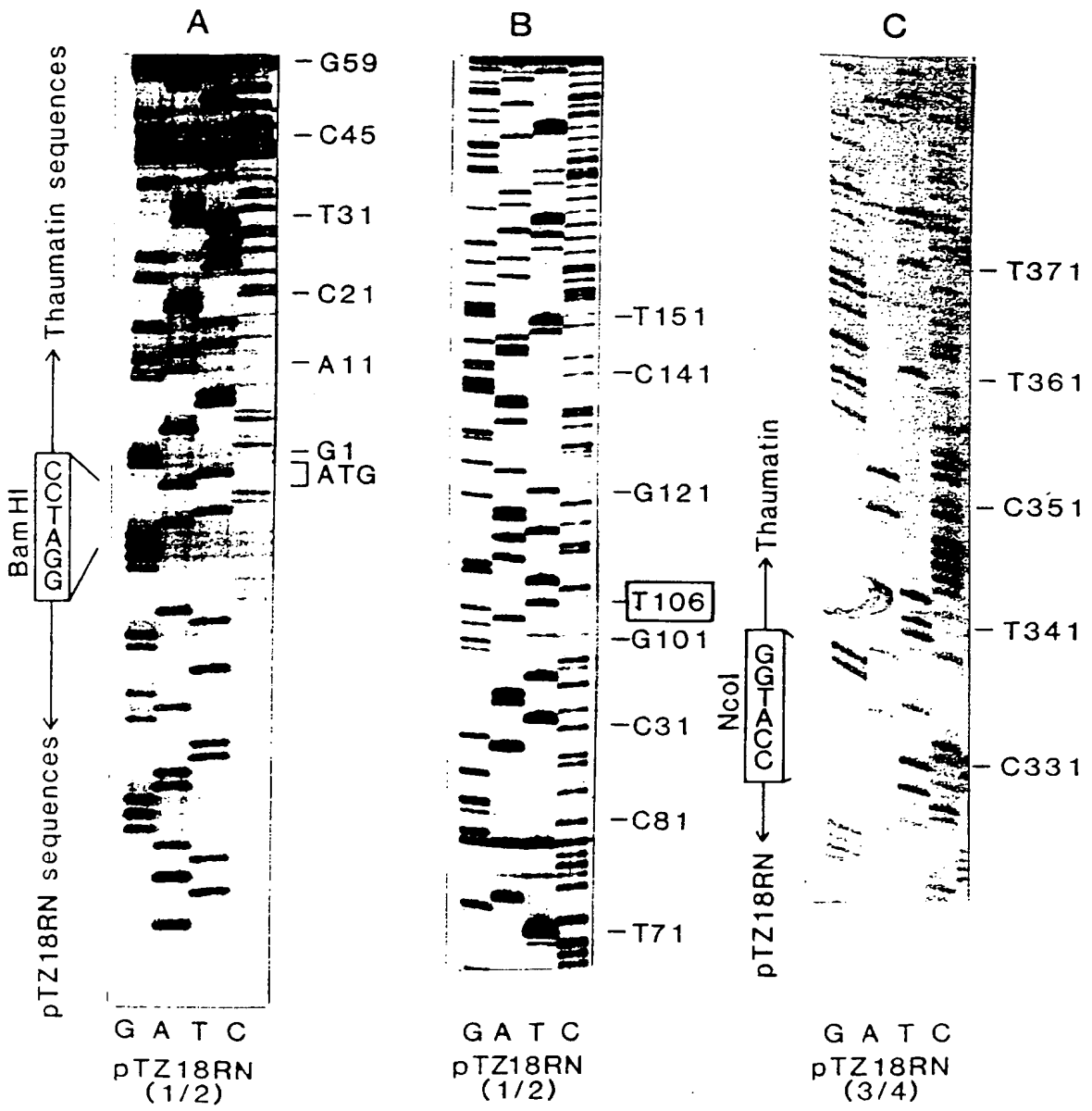


Fig. 5

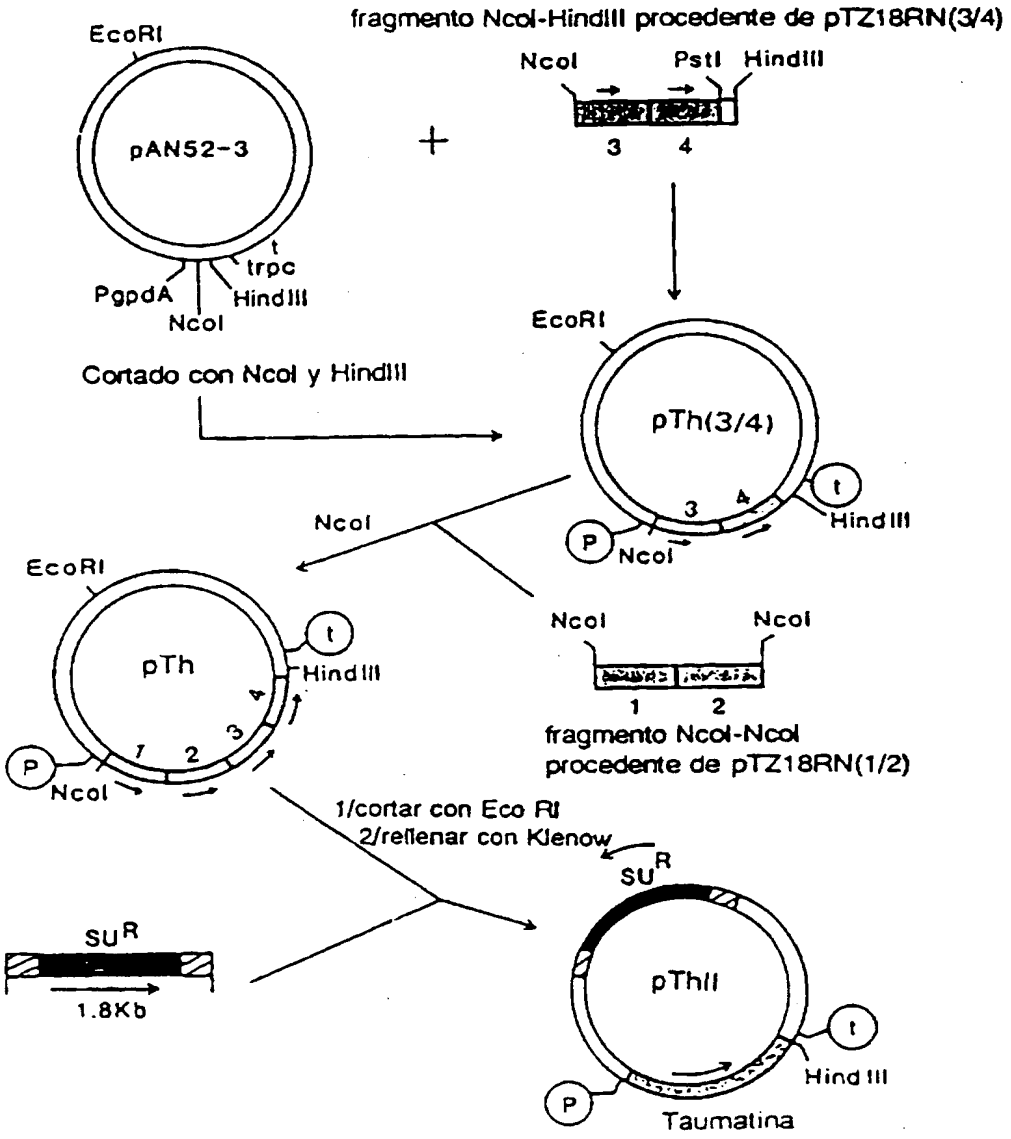


Fig. 6

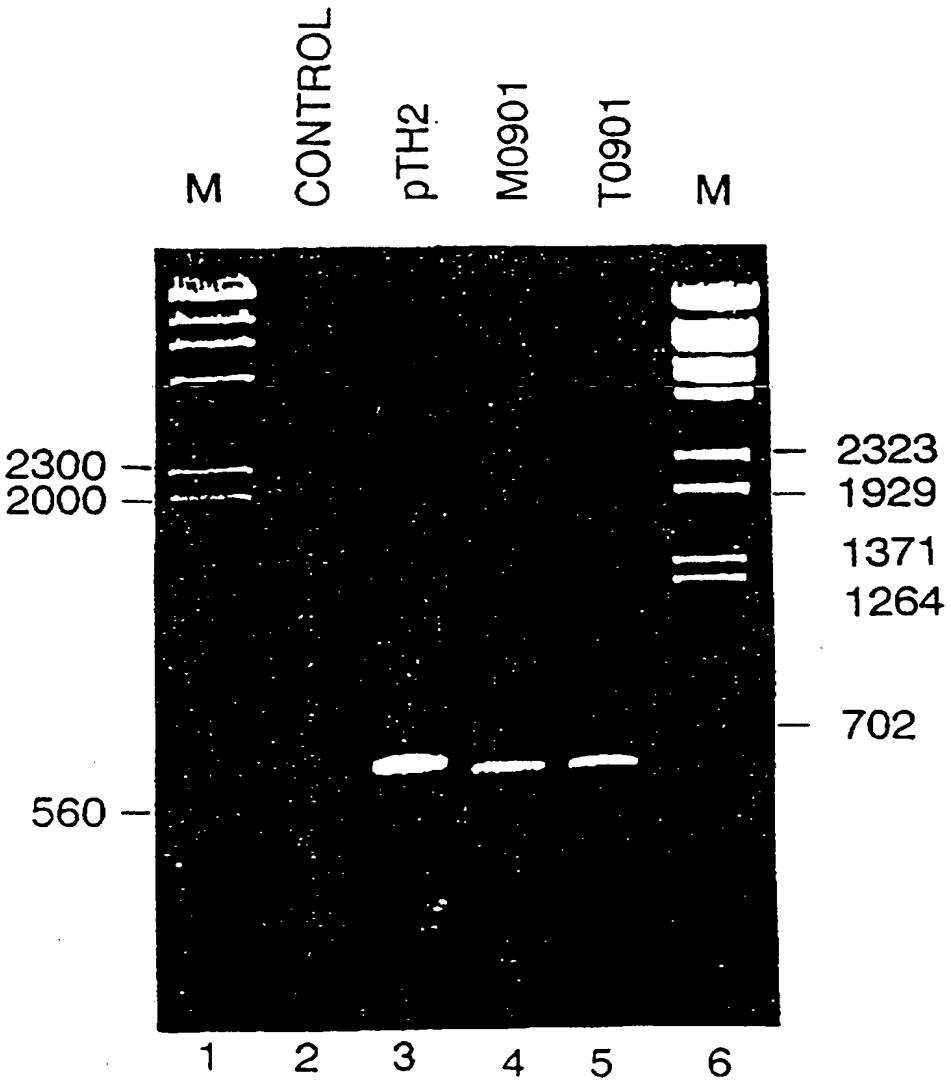

M
CONTROL
pTH2
M0901
T0901
M
2300
2000
560
2323
1929
1371
1264
702
1
2
3
4
5
6

Fig. 7

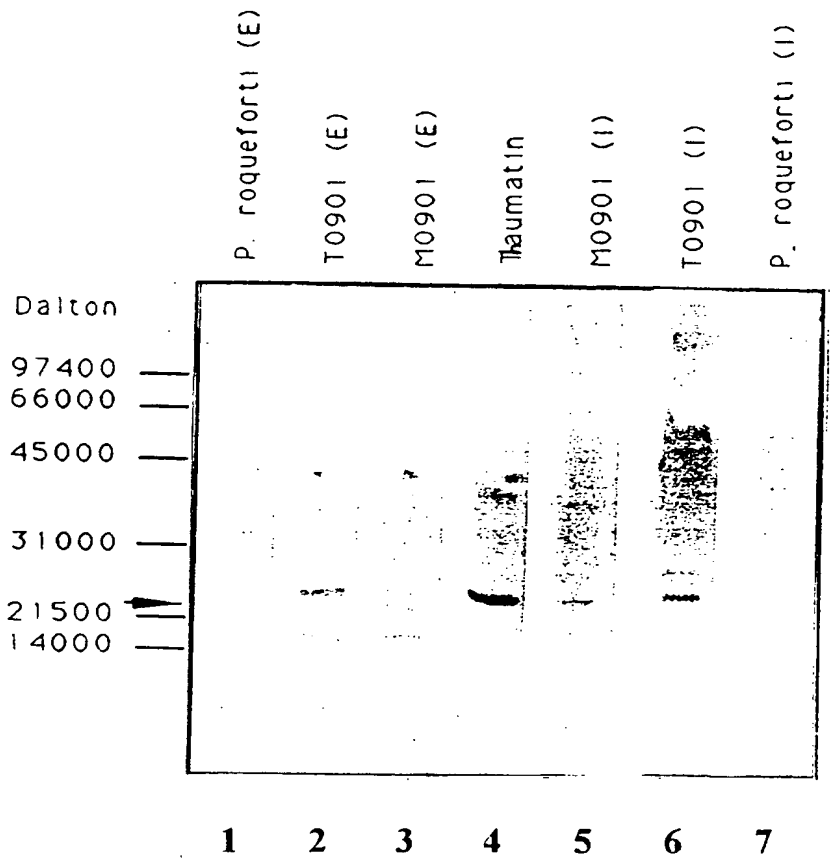

P. roqueforti (E)
T0901 (E)
M0901 (E)
Thaumatin
M0901 (I)
T0901 (I)
P. roqueforti (I)
Dalton
97400
66000
45000
31000
21500
14000
1
2
3
4
5
6
7

Fig. 8

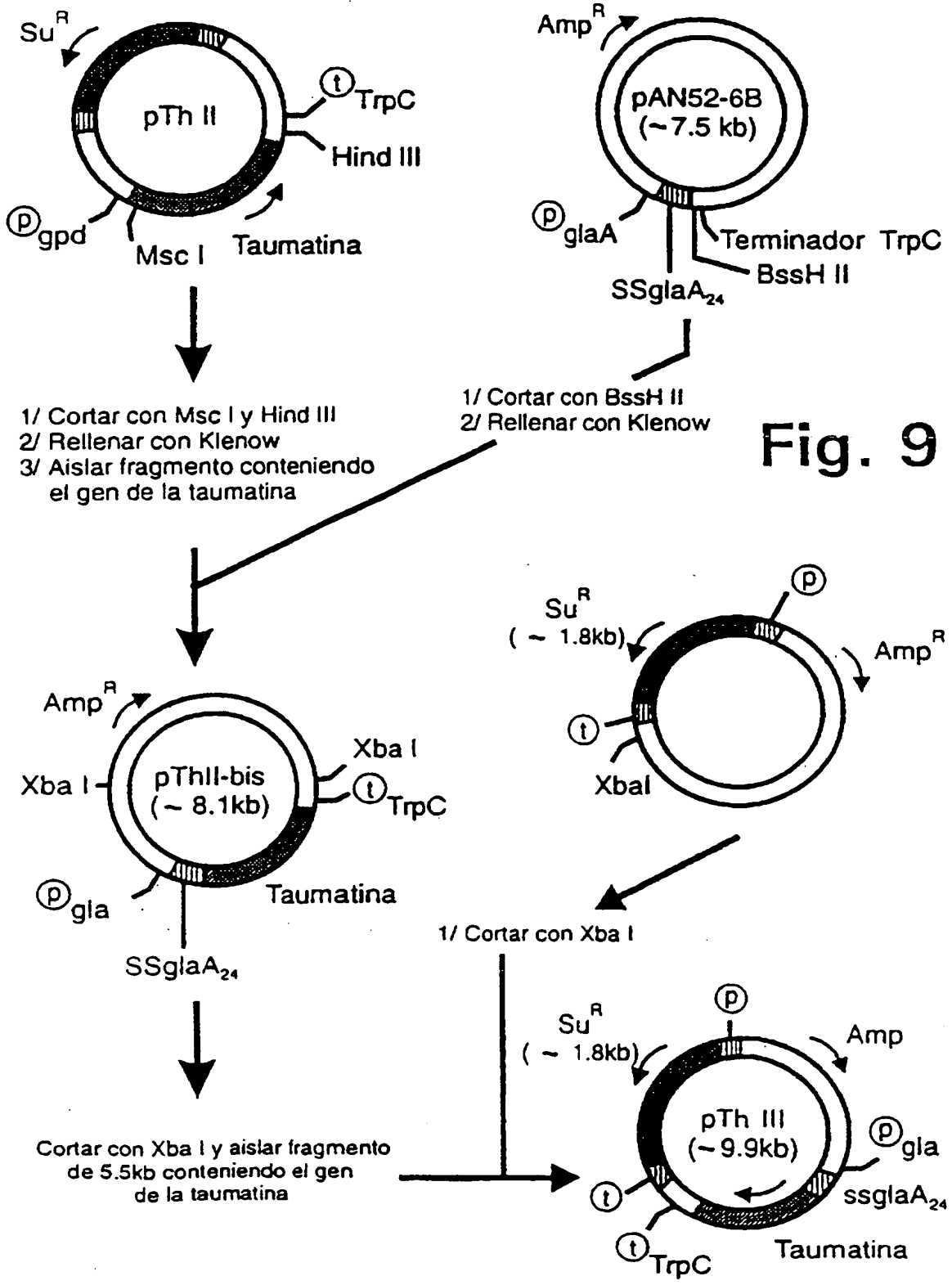


Fig. 9

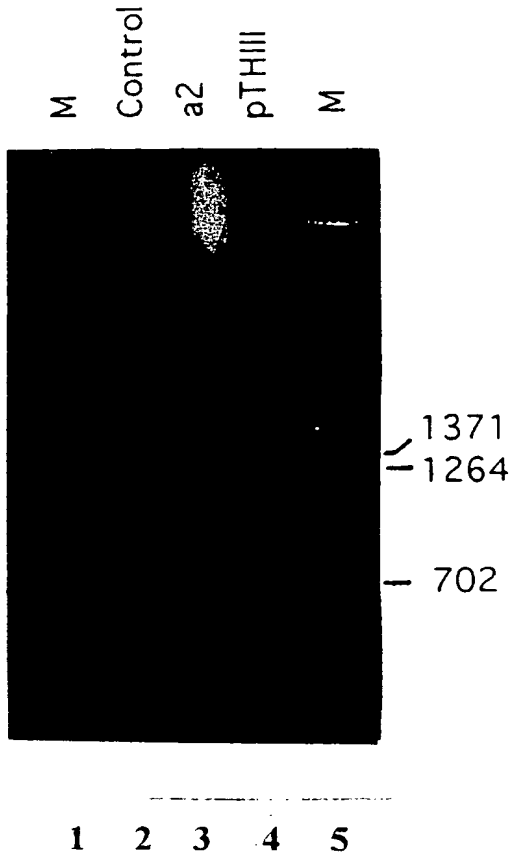


Fig. 10

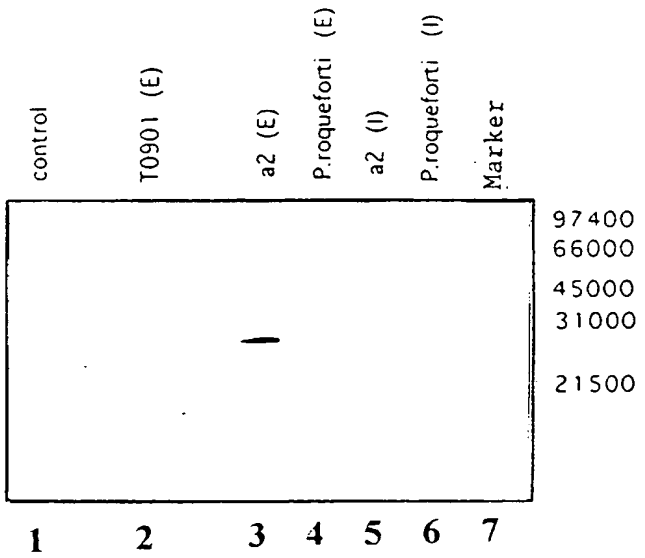


Fig. 11

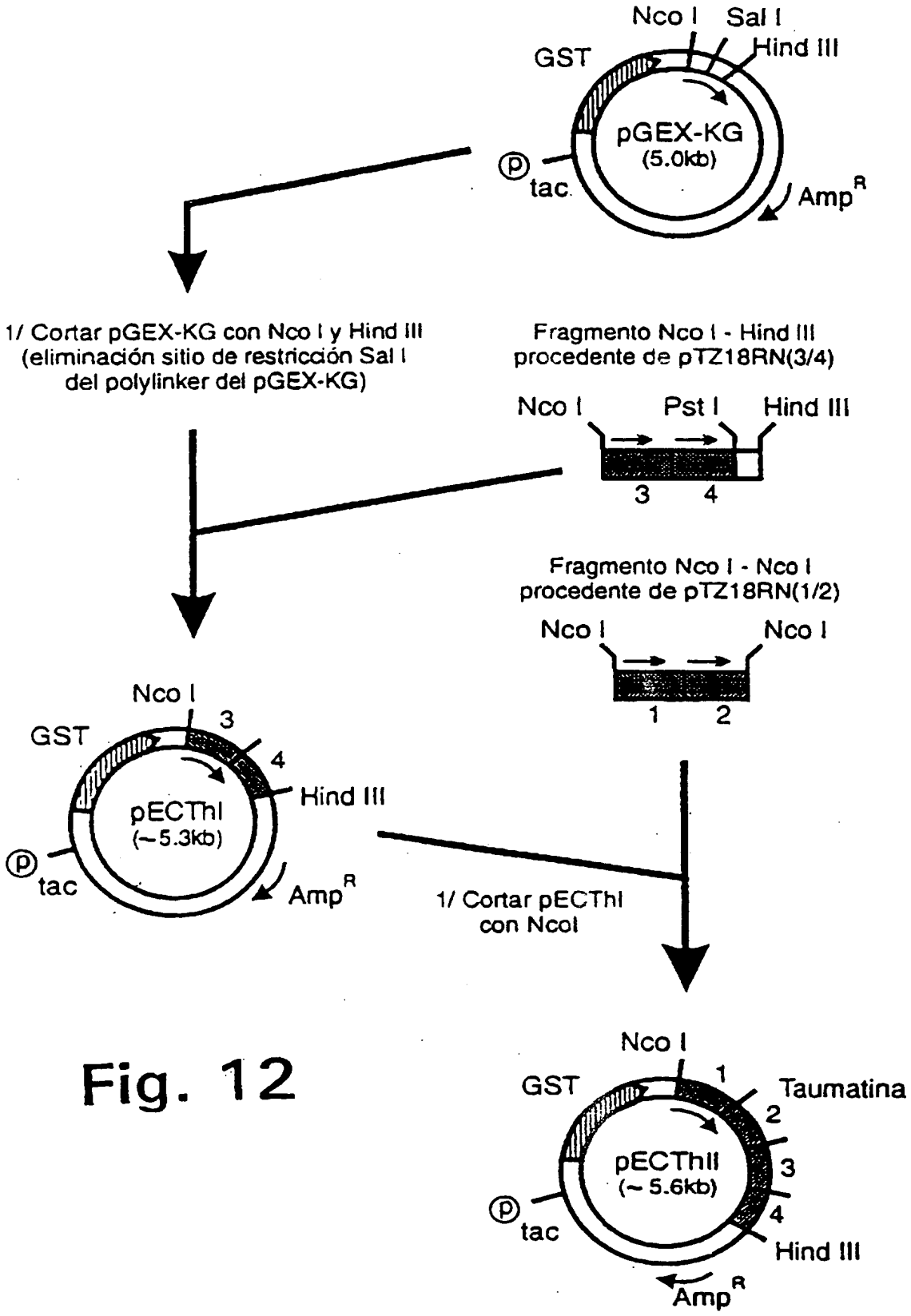


Fig. 12

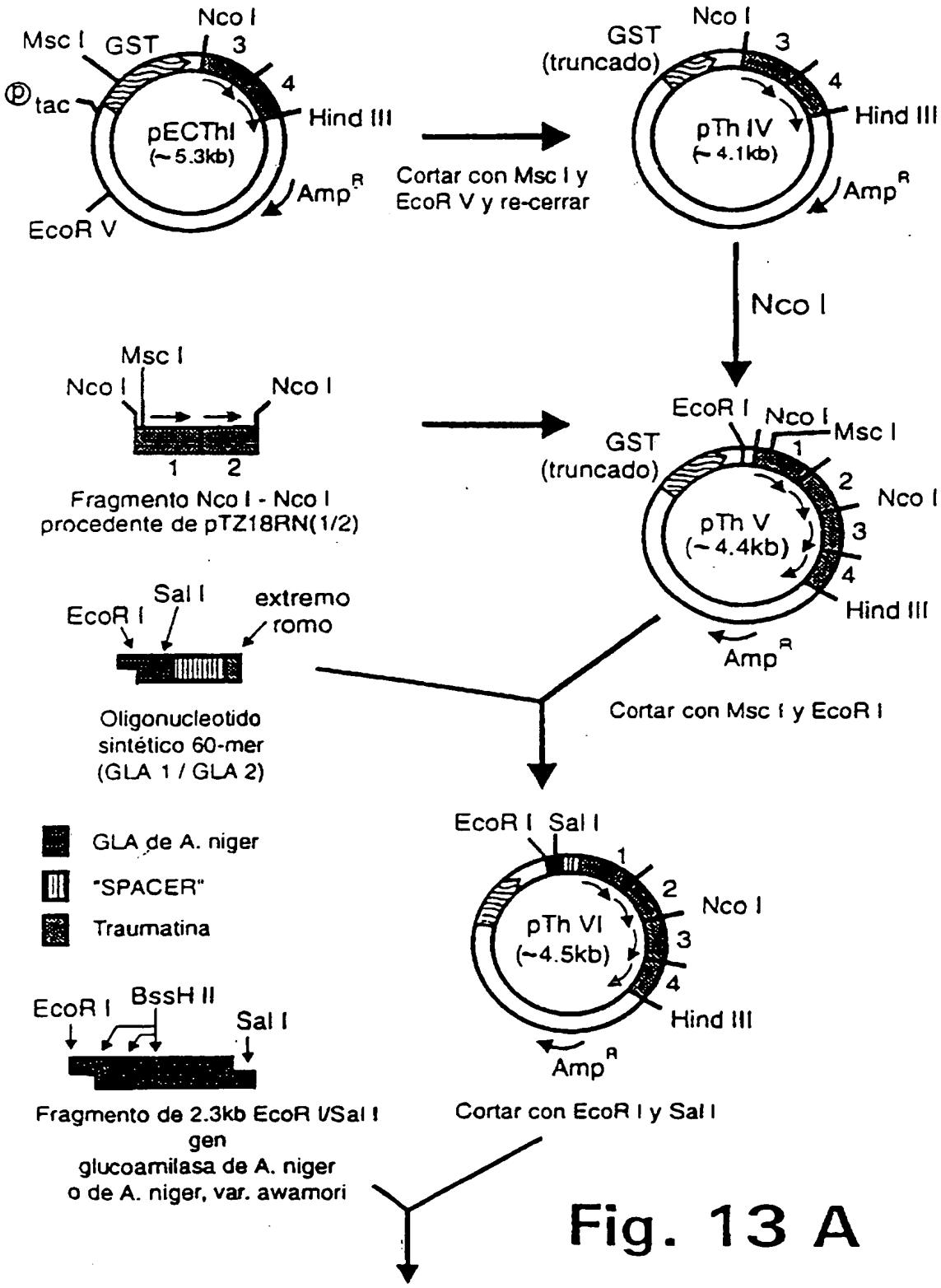
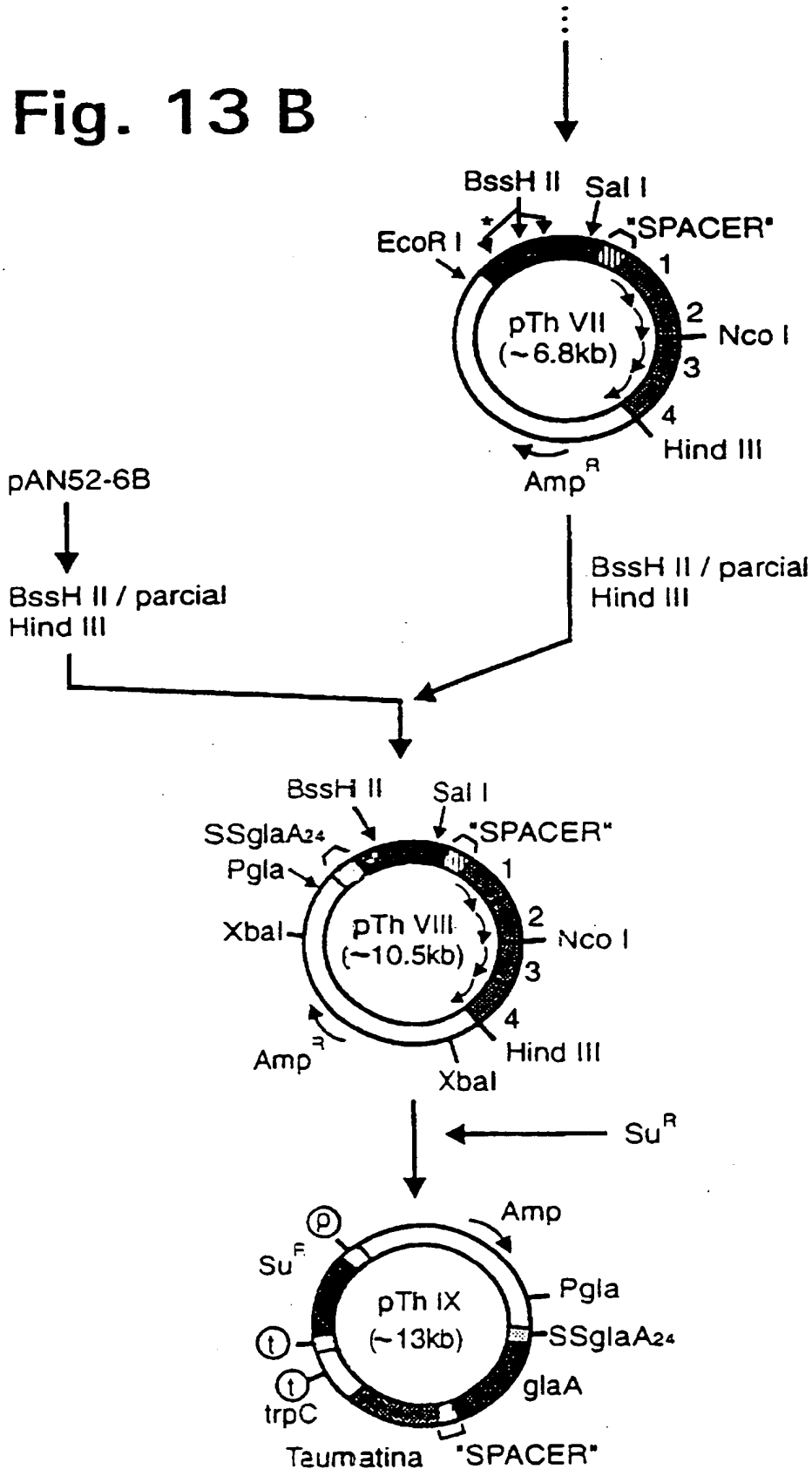


Fig. 13 A

Fig. 13 B



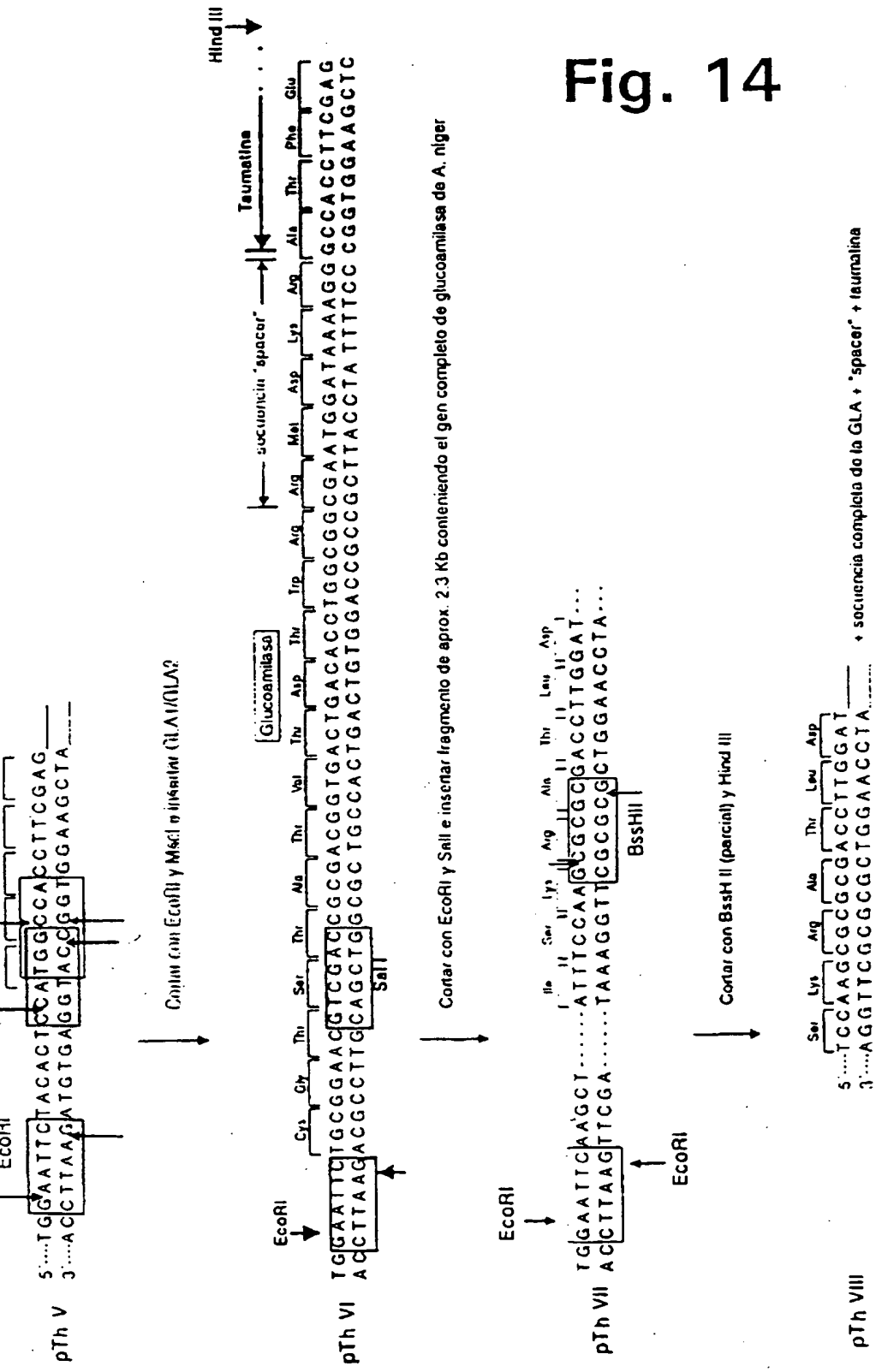


Fig. 14

(19) Europäisches Patentamt
European Patent Office
Office europeen des brevets

(11) EP 0 684 312 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
29.11.1995 Bulletin 1995/48

(21) Application number: **95105973.2**

(22) Date of filing: **21.04.1995**

(51) Int. Cl.[6]: **C12N 15/29**, C07K 14/43, C12N 15/62, C12N 1/15, C12N 15/80
// (C12N1/15, C12R1:66, 1:80)

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **21.04.1994 ES 9400836**

(71) Applicant: **URQUIMA S.A.**
E-08026 Barcelona (ES)

[72] Inventors:
- **Uriach-Marsal, Juan**
  E-08017 Barcelona (ES)
- **Rubio-Susan, Victor**
  E-28008 Madrid (ES)
- **Patino-Martin, Cristina**
  E-28760 Tres Cantos Madrid (ES)
- **Iossif Kalo-Koenova, Eliza**
  E-28028 Madrid (ES)
- **del-Moral-Juarez, Catalina**
  E-28031 Madrid (ES)
- **Faus-Santasusana, Ignacio**
  E-08035 Barcelona (ES)
- **del-Rio-Pericacho, José-Luis**
  E-08224 Tarrasa, Barcelona (ES)
- **Bladé-Piqué, Joan**
  E-08024 Barcelona (ES)

(74) Representative: **Zumstein, Fritz, Dr. et al**
Dr. F. Zumstein
Dipl.-Ing. F. Klingseisen
Bräuhausstrasse 4
D-80331 München (DE)

Remarks:
The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(54) Preparation process of a natural protein sweetener

(57) Thaumatin II or thaumatin I can be obtained through the expression, not of their natural genes, but of artificial, synthetic and substantially optimized genes following specific rules. Preferably, this expression is carried out in filamentous fungi, especially GRAS fungi and particularly the species Penicillium roquefortii, Aspergillus niger and the awamori variant of Aspergillus niger. Preparing substantially optimized artificial genes for filamentous fungi, performed here for the first time in the case of thaumatin, allows for high protein expression, making the process useful for industrial production of this valuable sweetener. Thaumatins may be obtained extracellularly by using a plasmid with a secretion signal, and also intracellularly. The latter method can be used in animal feed without prior separation from the fungal mycelium.

CORRIGENDUM issued on 05.06.96 bibliography updates included

EP 0 684 312 A2

Printed by Rank Xerox (UK) Business Services
2.12.4/3.4

Description

This invention is based on genetic engineering or recombinant DNA technology and refers to a process for obtaining natural proteinaceous sweeteners of the thaumatin type, to new DNA sequences which have been optimized for expression in filamentous fungi and which codify these proteins, and to the use of these sequences in the transformation of filamentous fungi for the production of thaumatins.

STATE OF THE ART

The thaumatins are proteins with a very sweet taste and the capacity to increase the palatability (upgrading or improving other flavours) of food; in industry they are currently extracted from the arils of the fruit of the plant Thaumatoccocus daniellii Benth. Thaumatins can be isolated from these arils in at least five different forms (I, II, III, b and c), which can be separated using ion-exchange chromatography. These forms are all single-chain polypeptides with 207 amino acids and a molecular weight of approximately 22,000 Daltons. Thaumatins I and II, which predominate in the arils and have very similar sequences of amino acids, are much sweeter than saccharose (100,000 times sweeter according to one estimate). Besides being natural products, thaumatins I and II are non-toxic, making them a good substitute for common sweeteners in the animal and human food industries.

Despite its advantages, industrial use of thaumatins of natural plant origin is very limited because of the extreme difficulty involved in obtaining the fruit from which it is extracted. The producing plant, T. daniellii, not only requires a tropical climate and pollination by insects, but it must also be cultivated among other trees and yet 75% of its flowers do not bear fruit.

Although attempts have been made to produce thaumatins by genetic engineering in bacteria such as Escherichia coli (cf. EP 54.330, EP 54.331 and WO 89/06283), Bacillus subtilis and Streptomyces lividans, in yeasts such as Saccharomyces cerevisiae (cf. WO 87/03007) and Kluyveromyces lactis (EP 96.430 and EP 96.910), in the fungus Aspergillus oryzae (Hahm and Batt, Agric. Biol. Chem. 1990, vol. 54, pp. 2513-20), and in transgenic plants such as Solanum tuberosum, until now the results have been considered disheartening; thus the thaumatin available to industry is very scarce and expensive (cf. M. Witty and W.J. Harvey, "Sensory evaluation of transgenic Solanum tuberosum producing r-thaumatin II", New Zealand Journal of Crop and Horticultural Science, 1990, vol. 18, pp. 77-80, and the articles cited therein).

Accordingly, there has remained a need for economically obtaining industrial amounts of thaumatins.

DESCRIPTION OF THE INVENTION

This invention solves the problem of preparing thaumatins II and I through their expression in filamentous fungi but without using natural DNA (or derived cDNA) as described for the fungus Aspergillus oryzae. Rather, artificial, synthetic and substantially optimized genes are used for expression in filamentous fungi according to specific rules.

Obtaining substantially optimized artificial genes for filamentous fungi, performed here for the first time for thaumatins, allows for high expressions of protein, making the process useful for industry.

In a specific embodiment of this invention, the filamentous fungi used belong to those considered innocuous, particularly to those included on the GRAS list (Generally Recognized as Safe). Preferred GRAS fungi include the Penicillium genus, especially the species Penicillium roquefortii, or the Aspergillus genus, especially the niger species and the niger variant awamori.

This invention encompasses obtaining thaumatins I and II secreted or produced extracellularly (for which an appropriate secretion signal must be introduced in the plasmid), and obtaining thaumatins I and II intracellularly, which allows for their use in animal food, without prior separation of the mycelium from the fungi.

The following abbreviations are used below, among others:

A	= Adenine
Amp	= Ampicillin
ATP	= Adenosine triphosphate
BSA	= Bovine serum albumin
C	= Cytosine
CIP	= Calf intestinal phosphatase
dATP	= 2'-Deoxyadenosine triphosphate
dCTP	= 2'-Deoxycytidine triphosphate
dGTP	= 2'-Deoxyguanosine triphosphate
DNA	= deoxyribonucleic acid
DTT	= 1,4-Dithiothreitol
dTTP	= 2'-Deoxythymidine triphosphate

EDTA	= Ethylenediaminetetra-acetic acid (disodium salt)
G	= Guanine
GRAS	= Generally regarded as safe
KDa	= Kilodalton
MCS	= Multiple cloning site
nt	= Nucleotides
bp	= base pairs
PCR	= Polymerase chain reaction
PEG	= Polyethylene glycol
PMSF	= Phenylmethylsulfonyl fluoride
rpm	= revolutions per minute
SDS	= Sodium dodecyl sulphate
SSC	= Sodium sodium citrate (0.15M NaCl; 0.015M sodium citrate)
T	= Thymine
TE	= Buffer 10 mM Tris-HCl, pH 8.0; 1 mM EDTA
U	= Units
X-gal	= 5-bromo-4-chloro-3-indo-β-D-galactose

Amino acids are designated by their standard abbreviations. For plasmids, the published notation in each case is used.

One part of the subject-matter of this invention is a gene for codifying thaumatin II which is artificial, synthetic and more than 50% optimized for expression in filamentous fungi; this gene consists of a DNA sequence which codifies the sequence of amino acids of Sequence ID No. 1 (corresponding to the 207 amino acids of the protein thaumatin II), followed by n stop sequences, where integer n is greater than or equal to 1; this DNA sequence is the result of making more than 50% of the possible modifications of the DNA sequence of the natural gene which codifies the 207 amino acids of thaumatin II (gene described in the literature and also included in Sequence ID No. 1) through the addition of one or more (n in Sequence ID No. 1) stop codons and performing more than 50% of the possible changes on the nucleotide codons corresponding to the thaumatin II amino acids; these changes consist of substituting the original codons in a given amino acid with the codon in parentheses in the following list of amino acid codons:
Ala (GCC), Arg (CGC), Asn (AAC), Asp (GAC), Cys (TGC), Lys (AAG), Gln (CAG), Glu (GAG), Gly (GGC), Ile (ATC), Leu (CTC), Met (ATG), Phe (TTC), Pro (CCC), Ser (TCC), Thr (ACC), Trp (TGG), Tyr (TAC), Val (GTC);
As is well known in the art, TAA, TAG or TGA can be used as stop codons, or any combination thereof.

The specific case of the previous gene in which an optimization of more than 75% was performed is preferred. It is even more preferred when the optimization is maximum (100%), i.e., when the DNA sequence of the artificial gene is obtained from the Sequence ID No. 1 sequence by performing 100% of the all possible codon changes, which corresponds to Sequence ID No. 2. Also preferred are the previous genes where n is between 1 and 3.

Another part of the subject-matter of this invention is a gene for codifying thaumatin I which is artificial, synthetic and more than 50% optimized for its expression in filamentous fungi; this gene consists of a DNA sequence which codifies the sequence of amino acids corresponding to the 207 amino acids of the protein thaumatin I (sequence of 207 amino acids which differs from those of Sequence ID No. 1 in only five amino acids, i.e., 46-Asn, 63-Ser, 67-Lys, 76-Arg and 113-Asn); this optimized DNA sequence is obtained by leaving the following five codons unchanged: AAC (46-Asn), TCC (63-Ser), AAG (67-Lys), CGC (76-Arg)and AAC (113-Asn), by modifying the rest of the codons as described above for the DNA sequence of the thaumatin II gene, and by adding one or more stop codons, as described above. The gene which codifies thaumatin I and which is more than 75% optimized is particularly preferred. It is even more preferred when the optimization is maximum (100%). Artificial genes to which between one and three stop codons have been added are preferred.

Hereinafter, any gene optimized more than 50%, more than 75% or up to 100% is called without distinction a "substantially optimized gene".

Subject-matter of this invention are also the recombinant plasmids made up of: (i) a substantially optimized gene for obtaining thaumatin I or II, (ii) an expression cassette for filamentous fungi containing an appropriate promoter sequence and a terminating sequence for this type of fungi, (iii) an appropriate selection marker, and (iv) an optional secretion signal DNA sequence for producing the protein extracellularly.

Particularly preferred are recombinant plasmids characterized in that the promoter sequence of the expression cassette comes from the gene of the enzyme glyceraldehyde 3-phosphate dehydrogenase of Aspergillus nidulans; the terminating sequence of the expression cassette is the tryptophan C sequence of Aspergillus nidulans; and the selection marker is that of resistance to sulfanilamide. Also preferred are the recombinant analogue plasmids where the promoter sequence of the expression cassette comes from the gene of the enzyme glucoamylase of Aspergillus niger.

In a particular embodiment of this invention, the recombinant plasmids used express the fusion protein thaumatin-glucoamylase, and they are characterized by comprising: (i) an appropriate selection marker; (ii) a DNA sequence made up of (a) a substantially optimized gene for the expression of thaumatin I or II, (b) a spacer sequence which in turn

contains a KEX2 processing sequence, and (c) the complete gene of the glucoamylase of Aspergillus niger or the awamori (glaA) variant of Aspergillus niger; and (iii) the "pre" and "pro" signal sequences of the glaA gene.

Part of the subject-matter of this invention are also the cultures of filamentous fungi capable of producing the proteins thaumatin I or II, which have been transformed with any of the abovementioned plasmids. In particular, the filamentous fungi of the species Penicillium roquefortii, Aspergillus niger and the awamori variant of Aspergillus niger are preferred.

Part of the subject-matter of this invention are also the production processes for thaumatin I or II which include the following steps:

a) incorporation of a substantially optimized gene for the expression of thaumatin I or II, in an expression vector selected from those corresponding to the abovementioned plasmids using standard recombinant DNA technology techniques;

b) transformation of a strain of filamentous fungus with the previous expression vector;

c) culture of a filamentous fungus strain transformed in this way in the appropriate nutrient conditions to produce thaumatin I or II, either intracellularly, extracellularly or through both methods simultaneously, or in the form of the fusion protein thaumatin-glucoamylase;

d) depending on the case, separation and purification of thaumatin I or II alone, or separation of thaumatin I or II from the culture medium, together with the fungal mycelium.

In a preferred embodiment of these processes, the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger or the awamori variant of Aspergillus niger.

To obtain thaumatin II, pThII recombinant plasmids are preferred, which can be obtained through the method described in the examples and illustrated in Figure 6, which can be summarized as follows: a) starting with plasmid pTZ18RN(3/4), a fragment (3/4) of the DNA sequence of the substantially optimized gene which codifies thaumatin II is obtained; b) this fragment is ligated with plasmid pAN52-3, generating plasmid pTh(3/4); c) starting with plasmid pTZ18RN(1/2), the remaining fragment (1/2) of the DNA sequence of the substantially optimized gene which codifies thaumatin II is obtained; d) this fragment is ligated to plasmid pTh(3/4), generating plasmid pTh; e) a DNA fragment is inserted to provide resistance to sulfanilamide, Su^r, thus obtaining plasmid pThII (Figure 6). With this plasmid, thaumatin II is obtained intracellulary for the most part.

For the production of thaumatin II in a basically extracellular way in Penicillium roquefortii, pThIII plasmids are preferred, the preparation of which is described in Example 2 and is outlined in Figure 9. To prepare it in the awamori variant of Aspergillus niger, the process described in Example 3 is used.

To produce thaumatin II as a fusion protein with glucoamylase, the pECThII and pThIX plasmids can be used, preparation of which is described in the examples and outlined in Figures 12, 13A and 13B.

To produce thaumatin I, the recombinant plasmids obtained following methods analogous to those used to produce thaumatin II are used. Thus, for example, for intracellular production in Penicillium roquefortii, pThI plasmids are used which are obtained as follows: a) Starting with plasmid pTZ18RN(1/2), the fragment (1/2) of the substantially optimized gene sequence is obtained which codifies thaumatin II; b) this fragment is ligated to plasmid pTZ18RN(3/4) linearized with NcoI, thus generating plasmid PTZ18RN(Th); c) starting with plasmid pTZ18RN(Th) in single-stranded form and using site-directed mutagenesis techniques, the following changes are carried out on the sequence of the synthetic and artificial gene of thaumatin II, where the symbol -> joins the replaced (original) and the replacement (final) in this order:

AAG -> AAC (46-Lys -> 46-Asn)
CGC -> TCC (63-Arg -> 63-Ser)
CGC -> AAG (67-Arg -> 67-Lys)
CAG -> CGC (76-Gln -> 76-Arg)
GAC -> AAC (113-Asp -> 113-Asn)

This plasmid is called pTZ18RN(ThI); d) starting with plasmid PTZ18RN(ThI) a DNA fragment of the complete sequence of the substantially optimized gene which codifies thaumatin I is obtained; e) this fragment is ligated to plasmid pAN52-3, thus generating plasmid pTh'; f) a DNA fragment containing resistance to sulfanilamide, Su^R, is inserted, thus obtaining plasmid pThI.

In a specific embodiment of this invention, the plasmids are replicated and amplified in Escherichia coli.

When the filamentous fungus is of the GRAS type, the processes for isolating thaumatin I or II together with the fungal mycelium are particularly interesting. In these cases, a part of the subject-matter of this invention is also the use of mixtures of thaumatin I or II and fungal mycelium obtained in this way to increase the sweetness or palatability of animal food.

When it is necessary to obtain purified thaumatin I or II, it is particularly important for the expression vector to be a plasmid which also contains a secretion signal sequence in the DNA so that the filamentous fungus produces thaumatin I or II extracellularly.

In some cases the production of thaumatin I or II can be increased by obtaining the fusion protein with glucoamylase.

In specific embodiments of this invention, when obtaining the pThI and pThII plasmids, the promoter sequence of the expression cassette can come from any gene from the following enzymes of filamentous fungi: glyceraldehyde 3-phosphate dehydrogenase, β-glucoamylase, alcohol dehydrogenase, glucoamylase or α-amylase. Moreover, the terminating sequence of the expression cassette can be the sequence corresponding to the promoter sequence in question. Finally, the selection marker can be of the type which is resistent to sulfanilamide, oleomycin, hygromycin B, phleomycin or acetamide.

As shown in the examples, this invention makes it possible to obtain thaumatin I or II for industry with satisfactory phenotypical characteristics, and with high productivity, which represents a considerable advantage over the state of the art.

Moreover, because the fungus is harmless, the thaumatin can be administered together with the mycelium, a fact which saves time in the purification process and, therefore, represents a considerable additional advantage, especially for use in animal feed.

Without being limiting, the following detailed examples illustrate this invention. The culture of the fungus Penicillium roquefortii, which produces the thaumatin II obtained in Example 1, has been deposited in the Spanish Collection of Standard Cultures (Colección Española de Cultivos Tipo, CECT) of the Departamento de Microbiologia of the Facultad de Ciencias Biológicas of the University of Valencia, with number CECT 2972.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: (A) DNA sequence showing nucleotides 272-304 from the MCS of commercial plasmid pTZ18R. (B) Fragment of plasmid pTZ18RN, obtained from the former, showing its unique NcoI restriction site.

Figure 2: Strategy used to build the synthetic gene with two pairs of oligonucleotides. Each pair of oligonucleotides has a complementary area. A, B and C represent restriction enzymes necessary for cloning of the oligonucleotide pairs, once they are paired and elongated, on the pTZ18RN vector.

Figure 3: Sequences of the oligonucleotides used to build the gene.

Figure 4: Diagram of the different stages in the construction of the artificial and synthetic gene (sequence represented in black).

Figure 5: Representative autoradiographs of the gene sequence using the Sanger dideoxy method: (A) the first 60 nucleotides; (B) nucleotides 70-170; (C) nucleotides 330-370.

Figure 6: Diagram of the manipulations performed to obtain the pThII plasmid.

Figure 7: Results of the PCR analysis of the two transformed fungi, M0901 and T0901, compared with the pThII plasmid and an untransformed control fungus. On the y-axis, the number of bases according to two standard reference markers.

Figure 8: Results of the immunoblotting analysis of the transformed fungi from Figure 7, compared with commercial thaumatin II (supplied by Sigma Inc.) and an untransformed control fungus (E = extracellular protein; I = intracellular protein). The numbers on the y-axis correspond to protein markers of known molecular weight. The arrow indicates the place where the comercial thaumatin (4) and the recombinant thaumatin migrate (2, 3, 5 and 6).

Figure 9: Diagram of the manipulations performed to obtain plasmid pThIII. The sequence corresponding to the gene of resistance to sulfanilamide (Su^R) is shown as the dark crosshatched section and the sequence of thaumatin is shows as the lighter crosshatched section. The section with vertical lines shows the different fungal promoter and terminating sequences, as well as the "signal" sequence of 24 amino acids from the glucoamylase gene (labelled $SSGlaA_{24}$ in the figure).

Figure 10: Results of PCR analysis of the A2 transformed fungus (thaumatin secretor). On the x-axis, the number of bases according to standard reference markers. Lanes 1 and 5 correspond to markers, lane 2 contains DNA from an untransformed fungus (control), and lane 3 contains DNA from fungus a2. Lane 4 is a positive control (DNA from plasmid pThIII).

Figure 11: Results of the immunoblotting analysis of the transformed fungi T0901 and a2. Lane 1 contains commercial thaumatin supplied by Sigma, Inc. Lane 7 corresponds to protein markers of known molecular weight (the molecular weights of each protein are indicated next to each lane). Lane 2 contains the culture medium in which the T09011 fungus was grown, a producer of intracellular thaumatin. Lanes 3 and 4 contain the culture medium in which the a2 fungus was grown (extracellular producer) and an untransformed fungus (control). Lanes 5 and 6 contain mycelium from these two fungi, respectively.

Figure 12: Diagram of the manipulations performed to obtain the pECThII plasmid. The dark crosshatched section represents the synthetic gene of tharnmatin II.

Figures 13A and 13B: Diagram of the manipulations performed to obtain the pThIX plasmid. The dark crosshatched section is the glucoamylase (glaA) sequence of Aspergillus niger or the awamori variant of Aspergillus niger. The wavy line section represents the glutathione-S-transferase sequence of Escherichia coli. The synthetic gene codifying thaumatin II appears as the lighter grey crosshatched section and the spacer sequence is between the genes of thaumatin and glucoamylase with vertical lines.

Figure 14: Details of the sequences in the fusion area between glucoamylase and thaumatin.

EXAMPLES

EXAMPLE I: INTRACELLULAR PRODUCTION OF THAUMATIN II IN PENICILLIUM ROQUEFORTII

(1.1) Construction of the synthetic, artificial and completely optimized gene encoding thaumatin II.

(1.1.1) Optimization of the DNA sequence of thaumatin II.

Starting with the sequences of known amino acids and nucleotides in the bibliography for thaumatin II and its corresponding natural gene (cf. for example: EP 54.330), reproduced in Sequence ID No. 1, the sequence of optimized DNA of Sequence ID No. 2 was designed, which codifies the same protein and where n = 3 (it has 3 TAA stop codons). The optimized sequence of Sequence ID No. 2 was obtained by performing the maximum number of changes on the codons of Sequence ID No. 1, replacing the original codons with the codons indicated in parenthesis on the following list of amino acid codons, when the latter where different from the originals:
Ala (GCC), Arg (CGC), Asn (AAC), Asp (GAC), Cys (TGC), Lys (AAG), Gln (CAG), Glu (GAG), Gly (GGC), Ile (ATC), Leu (CTC), Met (ATG), Phe (TTC), Pro (CCC), Ser (TCC), Thr (ACC), Trp (TGG), Tyr (TAC), Val (GTC);

(1.1.2) Construction of the pTZ18RN recombinant plasmid using site-directed mutagenesis.

Before beginning assembly of the synthetic gene of thaumatin II, a single NcoI restriction site was inserted in the multiple cloning site (MCS) of the multifunctional plasmid pTZ18R (supplied by Pharmacia Inc.). In this way plasmid pTZ18RN was generated ("N" for NcoI), the restriction site of which is shown in Figure 1. The insertion of the NcoI restriction site was performed using the site-directed mutagenesis technique described below:
Oligonucleotide p115 (5'-ACCCGGGGATCCTCTCCATGGGACCTGCAGGCATGCA-3') was supplied by Ingenasa S.A. (Madrid, Spain). Using standard procedures (Maniatis et al., "Molecular cloning, a laboratory manual", Cold Spring Harbor Laboratory Press, 1989), this oligonucleotide was labeled at the 5' end by transferring ^{32}P from [gamma-^{32}P]ATP with polynucleotide kinase. pTZ18R, with its DNA in single-stranded form, was obtained by standard techniques and was hybridized with one picomol of oligonucleotide labelled with ^{32}P at the 5' end in a buffer containing 40 mM Tris.HCl, pH 7.5, 50 mM NaCl and 20 mM $MgCl_2$ (final volume 5 µL). The mixture was incubated at 65°C for five minutes and allowed to cool slowly (overnight) to room temperature. The following enzymes and reagents were then added to the 5 µL of this mixture: 1.5 µL of 10X solution B (200 mM Tris.HCl, pH 7.5; 100 mM $MgCl_2$; 50 mM DTT); 1 µL of 10 mM ATP; 4 µL of a mixture containing 2.5 mM of each of the 4 dNTPs (dATP, dGTP, dTTP, dCTP); 6.5 µL of water; 1 µL of T4 DNA polymerase (3 units/µL); and 1 µL of DNA ligase (6 units/µL). The reactions were incubated for 3 hours at room temperature and at the end of that time 1 µL of T4 DNA polymerase was added (3 units) and 1 µL of DNA ligase (6 units). The reactions were allowed to continue for 60 more minutes at 37°C.

Aliquots of 1.0 µL of each reaction were used to transform E. coli strain JM103. Various colonies grown in LB/ampicillin (100 µg/mL) dishes were replated in dishes with fresh medium and analyzed (LB = Luria broth, a culture medium with the following composition: 1% bacto-tryptone, 0.05% yeast extract, 170 mM NaCl, pH 7.0). To be able to identify the clones containing the desired mutation, the colonies were analyzed using the p115 oligonucleotide labelled with [gamma-^{32}P]ATP as a probe, as described below.

Candidate colonies were replated in nitrocellulose filters (Schleicher & Schuell). The filters were placed in LB/amp dishes and incubated overnight at 37°C. The next day the cells were lysed by successively washing the filters in three solutions:

- Five minutes in 0.5 M Tris.HCl, pH 7.5, 1 M NaCl.
- Five minutes in 1 M Tris.HCl, pH 7.5.
- Five minutes in 0.5 M Tris.HCl, pH 7.5, 1 M NaCl.

The filters were then dried at 80°C for 90 minutes. Once they were dry, the filters were washed three times in 3X SSC, 0.1% SDS. Pre-hybridization took place in a solution containing 6X SSC, 5X Denhardt solution, 0.05% sodium pyrophosphate, 100 µg/ml of boiled salmon sperm DNA, and 0.5% SDS. Filters were pre-hybridized for one hour at

37°C. Hybridization took place overnight in 50 mL of the same solution, to which 33 ng of labelled p115 probe was added. The hybridization temperature was 50°C. On the next day the filters were washed as follows.

- First wash: 15 minutes in 2X SSC, 0.1% SDS, at room temperature.
- Second wash: the same conditions, but at 55°C.
- Third wash: The same conditions, but at 65°C.
- Fourth wash: 15 minutes in 0.4X SSC, 0.1% SDS at 65°C.

After the fourth wash, the filters were exposed to an X-ray film for 2 hours at -20°C. Various colonies with DNA showing marked hybridization to probe 115 were identified and DNA was extracted from each one.

The final identity of the clones was verified by testing if the DNA could be cut or not cut with NcoI and by analyzing its sequence. The plasmid containing the NcoI restriction site between the BamHI and PstI restriction sites (Figure 1) was called pTZ18RN and was the parent vector used in the construction of the artificial, synthetic and totally optimized gene of thaumatin II.

(1.1.3) Strategy for building the synthetic gene which codifies thaumatin II

The method chosen for assembling the synthetic gene of thaumatin II is shown in Figure 2. The eight long oligonucleotides whose sequences are shown in Figure 3 were supplied by Isogen Bioscience, Inc. (Netherlands). The single-stranded oligonucleotides, which occur in pairs, can be paired because of the complementary nature of the sequences. They were labelled 1a, 1b; 2a, 2b; 3a, 3b; and 4a, 4b. After pairing, the single-stranded areas were filled with the modified T7 DNA polymerase (the Taq DNA polymerase can also be used). The resulting double-stranded fragments were digested with the appropriate restriction enzymes to obtain cohesive ends or blunt ends and then ligated to the desired vector.

Figure 4 shows the strategy used to build the synthetic gene in 2 fragments which were then joined to an expression vector.

(1.1.3.1) Assembly of the first 332 pairs of bases of the synthetic gene of ID Sequence No. 2 (n = 3).

In the first stage, the oligonucleotides 1a, 1b, 2a and 2b were joined to obtain a DNA fragment with 332 base pairs which could be inserted into the pTZ18RN plasmid.

One microgram of oligonucleotide 1a and 1 µg of 1b were mixed in a buffer solution containing 40 mM Tris.HCl, pH 8.0, 10mM $MgCl_2$, 5mM DTT, 50 mM NaCl and 50 µg/mL of bovine serum albumin (BSA). The mixture (17 µL) was heated for 5 minutes at 70°C and then cooled slowly to 65°C for about ten minutes (appropriate temperature for hybridizing the pairs of oligonucleotides). Then 2 µL of a mixture of the four deoxynucleotides was added (2.5 mM of each dNTP) and 1 µL of the modified T7 DNA polymerase enzyme (Sequenase brand from U.S. Biochemical Corp.), giving a final volume of 20 µL. The reactions took place for 30 minutes at 37°C, followed by 10 additional minutes at 70°C (to inactivate the Sequenase). The reaction products were digested with Bam HI and Bgl II at 37°C for 3 hours. The following extractions were performed on the DNAs: once with phenol, once with phenol:chloroform and once with chloroform; they were then precipitated with ethanol. They were finally frozen in TE buffer at -20°C until later use.

The 2a and 2b oligonucleotides were processed in the same way except that the final products were digested with Bgl II and Nco I.

Plasmid pTZ18RN was digested sequentially with Bam HI and Nco I and was dephosphorylated with calf intestinal phosphatase (CIP). The linearized fragment of 2871 pairs of bases was recovered from a 0.8% agarose gel and then purified.

Then the products of reactions 1 and 2 were joined with the linearized pTZ18RN and the mixture was used to transform E. coli strain NM522. To identify the clones with the insert, a white/blue indicator test was used which works basically as follows:

The pTZ18R plasmid and its derivative pTZ18RN contain the bacterial gene LacZ'. Therefore, the bacterial colonies containing this plasmid are blue on dishes with LB/ampicillin which also contain the chromogenic substrate 5-Bromo-4-chloro-3-indo-β-D-galactose (X-gal). When a fragment of foreign DNA is inserted in the multiple cloning site (MCS) of the pTZ18RN plasmid, the LacZ' gene is deactivated and the resulting colonies are not blue, but white. Therefore, the white colonies were initially isolated, given that they were candidates for containing the different fragments of the synthetic gene of thaumatin II.

Various colonies with inserts of the appropriate size contained complete fragments of the 325 base pairs of the synthetic gene of thaumatin II. The resulting plasmid was called pTZ18RN(1/2).

(1.1.3.2) Assembly of the second 305 pairs of bases of the synthetic gene of ID Sequence No. 2 (n = 3)

In this case, an alternative approach was put into practice using Taq DNA polymerase and the PCR technique.

Before the annealing stage, oligonucleotides 3b and 4a were labelled at their 5' ends with a phosphate group using standard techniques. The oligonucleotides were called 3b* and 4a*.

One microgram of 3a and 1 µg of 3b* were incubated in a reaction mix (18 µL) containing 10 mM Tris.HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.1 mg/ml of gelatin. The samples were incubated for 5 minutes at 70°C and for five more minutes at 65°C. At this point, each dNTP was added (G, A, T, C) at a final concentration of 2 mM and 2.5 units of Ampli Taq DNA polymerase (Perkin-Elmer Cetus). The PCRs were as follows: 1 minute at 94°C; 1 minute at 55°C; and 1 minute at 72°C for 30 cycles, followed by a final extension at 72°C for 5 minutes. The samples were then extracted with phenol:chloroform and resuspended in 10 µL of TE buffer and incubated with Nco I at 37°C for 3 hours. After extracting and precipitating with ethanol, the DNAs were dissolved in TE buffer and frozen at -20°C until later use.

The 4a* and 4b oligonucleotides were processed as described above, except that the final products were digested with Pst I.

Ligation of the three fragments was done as per the same process mentioned above, except that pTZ18RN was used which was cut with Nco I and Pst I, treated with calf intestinal phosphatase and finally purified from an agarose gel. The ligation reactions contained 15% polyethylene glycol (PEG), which stimulates ligations with blunt ends. The ligation products are used to transform E. coli NM 522. A white/blue selection was made again of the recombinants on dishes with LB/amp medium supplemented with X-gal and IPTG. After analyzing the transformants, one clone was isolated which contained the 305 pb fragment of the second part of the thaumatin II gene. This plasmid was called pTZ18RN (3/4).

(1.1.3.3) Sequence Analysis

The identity of the synthetic gene was verified by analyzing its sequence using the Sanger method (Sanger, F. et al., Proc. Nat. Acad. Sci. USA 1977, vol. 74, p. 5463-67). A sequentiation kit was used (version 2.0) from United States Biochemical Corp. The sequence of the synthetic gene was determined without ambiguity by: (1) sequentiation of the two gene strands; and (2) performing parallel sequentiation reactions with dITP to destabilize the potential secondary structures which could form due to the areas rich in GC. Representative autoradiographs are shown in Figure 5.

(1.2) Insertion of the gene in an expression vector for filamentous fungi (Figure 6)

In this example, the pAN52-3 plasmid (described in Punt, P. J. et al., Journal of Biotechnology, 1990, vol. 17, pp. 19-34; called "starting plasmid" hereinafter) was the starting plasmid for construction of the expression vector in filamentous fungi (pThII) used to transform Penicillium roquefortii. Ligating the synthetic gene to this starting plasmid was performed in three stages described below.

(1.2.1) Ligating the 3/4 fragment

Thirty micrograms of pTZ18RN(3/4) was cut sequentially with Nco I and Hind III, generating 2 fragments. The small fragment with 310 bp containing the second part of the synthetic gene was purified in a 2% agarose gel. At the same time, 5 µg of the starting plasmid was cut sequentially with Nco I and Hind III. It was then dephosphorylated with alkaline phosphatase and a 5.8 Kb fragment was isolated in a 0.8% agarose gel. Then the starting plasmid, cut with Nco I and Hind III, dephosphorylated and purified, was ligated with the fragment of 310 bp from pTZ18RN(3/4). The mixture was used to transform E. coli DH5αF' as outlined in Figure 6. The clones containing the desired construction were identified by cutting the recombinant plasmids pTh(3/4) with Nco I and Hind III.

(1.2.2) Ligating fragment 1/2

In a second stage, plasmid pTZ18RN(1/2) was cut with Nco I and a NcoI-NcoI fragment containing the first part of the gene was purified in a 4% agarose gel. Plasmid pTh(3/4) was linearized with Nco I and processed with alkaline phosphatase. It was then ligated with the NcoI-NcoI fragment from pTZ18RN(1/2). The resulting plasmid was called pTh.

To analyze the clones, the pTh plasmid was cut with Bal I and Hind III. In the clones with the appropriate orientation, a fragment of 625 bp was obtained while those with inappropriate orientation produced a fragment of 300 bp.

(1.2.3) Ligating with the fungal marker

The pTh plasmid was then cut with Eco RI and the 5' ends were filled with the Klenow fragment of DNA polymerase I. This treated plasmid was then purified in a 0.8% agarose gel.

Starting with plasmid pEcoliR388 (N Datta, Saint Mary's Hospital, London), the sequence of resistance to sulfanilamide was obtained and a construction was made eliminating the procaryote promotor and terminator; then the structural gene was placed under the control of a promotor and a terminator of filamentous fungi (TrpC). The sulfanilamide resistance sequence obtained in this way was cut with SmaI and XbaI; the 5' ends were filled with Klenow and dNTP and a 1.75 Kb fragment was isolated from a 4% agarose gel. Then the fragment obtained in this way was ligated with pTh and transformation was carried out in E. coli DH1. The resulting plasmid was called pThII. This plasmid contains: (i) the synthetic gene which codifies thaumatin II under the control of a fungal promotor, and (ii) a sulfanilamide resistance marker. The final identity of the plasmid was verified by sequentiation as described in section 1.3.3.

(1.3) Transformation of Penicillium roquefortii with the aforementioned fungal expression vector

(1.3.1) Protoplast preparation

The protoplasts of Penicillium roquefortii used in the transformation experiments were prepared according to the following process, starting with the MUCL 29148 strain. Its conidia were inoculated in 50 mL of MSDPM liquid medium (medium semi-defined for mycelium production, the composition of which is described below). The culture was incubated for 44 hours at 28°C in a mechanical stirrer at 270 rpm. The mycelium was recovered by filtration, washed with sterile water and resuspended in a 1.2M KCl solution containing 40 mg of Lysin Enzyme (Sigma) per gram of mycelium. After 4 hours of incubation at 28°C at moderate stirring speed, protoplasts were obtained. Cell debris was eliminated by glass wool filtration. The protoplast suspension was washed and centrifuged (2000 rpm, 10 min.) twice with a 1.2 M KCl solution (10 mL/g). Finally, the protoplasts were resuspended in 1.2 M KCl (1 mL/g). This protoplast suspension (10^7-10^8 prot/mL) was used for the transformation experiments.

(1.3.2) Transformation

The protoplasts were centrifuged (2000 rpm, 10 min.) and then resuspended (5 x 10^8 protoplasts/mL) in solution I: 1.2 M KCl; 50 mM Tris.HCl (pH 8), 50 mM $CaCl_2$ and 20% of solution II (see below). They were incubated for 10 minutes at 28°C. Aliquots of 0.1 mL were mixed with DNA (10 µg) from the expression plasmid, which contained the thaumatin II gene. Immediately afterward, 2 mL of solution II [1.2 M KCl; 50 mM Tris.HCl (pH 8), 50 mM $CaCl_2$ and 30% PEG 6000] was added. This mixture was incubated for 5 minutes at room temperature. After recovering the protoplasts by centrifugation (2000 rpm, 10 min.), they were resuspended in 1 mL of 1.2 M KCl. Finally, aliquots of the protoplasts treated in this way were replated in petri dishes containing an appropriate medium for regeneration of the cell wall and subsequent selection using sulfanilamide (750 µg/mL). Using this transformation method, various strains that are resistant to sulfanilamide were isolated. These strains were analyzed to verify if the synthetic gene of thaumatin II had been incorporated into its genome.

(1.4) Analysis of the transformants

(1.4.1) PCR analysis

Analysis of the transformants obtained as described above to detect the DNA sequences of the synthetic gene of thaumatin II and resistance to sulfanilamide was performed using standard PCR techniques with appropriate oligonucleotides. Specifically, the T1 and T2 oligonucleotides were used, the sequences of which are included in section (1.4.1.2). T1 is complementary to nucleotides 605 and 624 of the upper strand of the synthetic gene of thaumatin II, while T2 is complementary to nucleotides 21 to 46 of the lower strand. Therefore, with these two oligonucleotides it was possible to amplify a fragment of 604 pairs of bases corresponding to oligonucleotides 21 to 624 of the synthetic gene of thaumatin II.

Figure 7 shows the success of the results, indicating that in the untransformed fungus (control), no bands appear of the size corresponding to the synthetic gene (lane 2), while in two of the transformant genes (M0901 and T0901) bands appear with the same number of bases as the band corresponding to the synthetic gene inserted in the pThII plasmid (lanes 3 to 5).

(1.4.1.1) Extraction of nucleic acids

The starting material was 5 g of mycelium which had been vacuum filtered using a Büchner funnel and which came from a 5-day MSDPM culture (0.6% $NaNo_3$; 0.052% $MgSO_4 \cdot 7H_2O$; 0.052 KCl; 1% glucose; 0.5% yeast extract; 0.5% casamino acids; $FeSO_4 \cdot 7H_2O$ traces; $ZnSO_4 \cdot 7H_2O$ traces).

The mycelium was ground in liquid nitrogen with a porcelain mortar. The mycelium was resuspended in the extraction buffer (10 mM Hepes, pH 6.9; 0.3 M saccharose; 20 mM EDTA, pH 8.0; 0.5% SDS) at a ratio of 10 mL of buffer per gram

of mycelium. It was incubated for 15 minutes at 65°C and centrifuged for 5 minutes at 7000 rpm (Beckman JA20 rotor) at room temperature to eliminate cell debris; the supernatant was collected and treated twice with phenol/chloroform/isoamyl alcohol (49:49:2) to eliminate proteins. The aqueous phase was precipitated with 0.3 M sodium acetate and 2.5 volumes of ethanol for 20 minutes at -20°C. The precipitated volume was centrifuged at 7000 rpm for 20 minutes. The precipitate was resuspended in 1 mL of TE buffer, pH 8.0.

(1.4.1.2) PCR reaction mix

In a total volume of 100 µL, 20 ng of DNA and 10 µL of PEC 10X buffer were mixed (500 mM KCl; 15 mM $MgCl_2$; 100 mM Tris HCl, pH 8.3; 0.01% porcine gelatin; a mixture of DNTPs, with a concentration of 200 µM of each; 2.5 units of Amplitaq and 1 µM of primer). The synthetic oligonucleotides used were T1 (26 nucleotides) and T2 (20 nucleotides) and specific primers for the beginning and end of the synthetic gene of thaumatin II.

T1: 5'-CCGCTGCTCCTACACCGTCTGGGCCG-3'

T2: 5'-TTAGGCGGTGGGGCAGAAGG-3'

Twenty µL of mineral oil was added to the mixture to keep the sample from evaporating.

(1.4.1.3) PCR

The sample underwent a cycle at 94°C for 5 minutes to separate the two DNA strands. Thirty chain reactions were then performed: first the DNA was denatured for 1 minute at 94°C; the temperature was lowered to 55°C for 30 seconds to allow the specific primers to join with the denatured DNA strand; the temperature was then increased again to 72°C for 1 minute to allow the new strand (in formation) to elongate. When all the cycles were completed, a final elongation was performed for 5 minutes at 72°C. The products of each PCR were analyzed in 0.8% agarose gel (Figure 7). Using this method two strains were identified called M0901 and T0901, the genomes of which contained the synthetic gene of thaumatin II.

(1.4.2) Immunoblotting Detection (Western-Blot)

Once the transformants that had incorporated themselves into the thaumatin II gene were detected correctly, Western blot was performed on the expression (Burnette W.N., Analytical Biochemistry, 1981, vol. 112, pp. 195-203), using polyclonal antibodies which had been previously obtained through standard rabbit immunization techniques to identify the protein. The serum obtained from each rabbit was precipitated with ammonium sulphate using standard techniques to precipitate the immunoglobulins, thus producing a protein fraction enriched with IgG antibodies. Figure 8 shows the outcome of the results obtained, indicating that no bands of the size corresponding to thaumatin II appear in the untransformed fungus (control), while in two of the transformed fungi a band appears having the same molecular weight as commercial thaumatin II.

(1.4.2.1) Preparation of the samples

The starting material was 2 g of mycelium which had been vacuum filtered using a Büchner funnel and which came from a 5-day culture at 28°C in MSDPM medium. Both the mycelium retained in the funnel (solid fraction) and in the culture medium (liquid fraction) were analyzed.

Solid Fraction

Ten mL of sonication solution (625 mM Tris.HCl, pH 6.5, 1mM PMSF, 5% β-mercaptoethanol) per gram of mycelium was added to the mycelium retained in the funnel. The mycelium was sonicated for 1 minute with 1-second pulses (i.e., 1 second sonificated, 1 second without sonification, and so on). The process was repeated three more times at intervals of from 3 to 5 minutes. It was centrifuged at 7500 rpm (Beckman JA20 rotor) for 20 minutes at 4°C.

Liquid Fraction

β-Mercaptoethanol (final concentration 5%) and PMSF (final concentration 1 mM) were added to 3 mL of the extracellular medium. Three mL of both fractions was used to start and was concentrated by column centrifugation (Bio-Rad ultrafilters) which retain the proteins having a molecular weight greater than 10,000 Daltons. In this process, the 3 mL passing through the columns was reduced to 200 µL.

Twenty µL of the 2 x sample buffer (25% glycerol; 2.5% SDS; 0.25M Tris.HCl, pH 7.0; 10 mM EDTA, pH 8.0; 0.002% bromophenol blue) was added to 20 µL of the concentrated solutions. They were boiled for 5 minutes and immediately placed in protein denaturing gel (SDS-polyacrilamide).

The protein gels used were 14% polyacrilamide and 18% urea. Electrophoresis was performed at 150 volts and stopped when the front of the sample was 3 or 5 mm from the end of the gel.

(1.4.2.2) Transfer to nitrocellulose

Once the eletrophoresis was completed and after removing the piled-up part, the gel was transferred to nitrocellulose paper (NC). To do so, the Bio-Rad Trans-blot SD Semidry Unit was used. Transfer took 30 minutes at 15 volts.

Once the bands were transferred to NC paper, the paper was left in blocking solution (3% BSA; 0.01% sodium aside; 0.05% Tween-20 in TBS; TBS = 150 mM NaCl; 50 mM Tris.HCl, pH 8.0) and stirred overnight. After this operation, the NC paper was processed as follows.

The NC paper was taken out of the blocking solution, washed with TBS and incubated with serum: immune IgG fraction (0.37 mg/mL) diluted (1:500) in blocking solution (with sodium azide). As a negative control, the normal preimmune IgG fraction was used (0.35 mg/mL) diluted (1:500) in blocking solution (with sodium azide). The solution was stirred and incubated for 4 hours at room temperature.

Three 10-minute washes were performed in TBS-Tween (TBS 1X + Tween-20, 0.05%). It was stirred and incubated for 4 hours at room temperature with the secondary antibody: anti-rabbit IgG-phosphatase alkaline conjugate diluted (1:500) in blocking solution (without sodium azide). Three 10-minute washes were performed in TBS-Tween.

The alkaline phosphatase reaction was performed: a) the NC was equilibrated with alkaline phosphatase buffer (100 mM Tris.HCl, pH 9.5 100; 100mM NaCl; 50 mM $MgCl_2$); b) the NC was placed in the development reaction mix (15 mL of alkaline phosphatase buffer, 66 µL of nitro blue tetrazodium, NBT) (75 mg/mL in 70% dimethyl formamide), 99 µL of 5-bromo-4-chloro-3-indole phosphate (BCIP) (25 mg/mL in 100% dimethyl formamide) until the bands turned dark; c) the reaction was stopped with alkaline phosphate stop solution (20 mM Tris.HCl, pH 8.0 and 20 mM EDTa, pH 8.0).

(1.4.2.3) Protein gel staining

The gels were stained for 1 hour with staining solution and stirred gently (25 ethanol; 10% acetic acid; 0.1% Comassie blue). They were destained with destaining solution (25% methanol; 7.5% acetic acid) until the blue color faded from the gel base.

EXAMPLE 2: EXTRACELLULAR PRODUCTION OF THAUMATIN IN PENICILLIUM ROQUEFORTII

For extracellular production of thaumatin, Penicillium roquefortii was transformed with plasmid pThIII, which was constructed as described below and outlined in Figure 9.

Plasmid pThII, described above, (section 1.2.3) was purified using standard techniques and resuspended in TE buffer at a final concentration of 1 µg/µl. Thirty micrograms (µg) of this plasmid were cut with restriction enzymes MscI and HindIII, and a fragment of 646 base pairs containing the complete gene of thaumatin II was isolated in a 0.8% agarose gel. The ends of the fragment were converted to blunt ends with the Klenow fragment from DNA polymerase I.

Plasmid pAN52-6B, containing approximately 7.5 Kb and derived from pAN52-6 Not 1 (cf. Van den Hondel et al., "Heterologous Gene Expression in filamentous fungi"; in Bennett and Lasvre, "More Gene Manipulation in Fungi"; Academic Press, 1991, chapter 18, pp. 396-428) was digested with BssHII and its ends were converted to blunt ends through the action of the Klenow fragment of DNA polymerase I.

These two fragments were ligated using DNA ligase and the resulting mix was used to transform the DH5αF' strain of E. coli. The resulting plasmid, pThII-bis, was isolated and its structure verified by sequencing using the Sanger dideoxy method.

The following step was to cut the pThII-bis plasmid (8.1 Kb) with XbaI and to isolate a fragment of approximately 5.5 Kb in length containing the thaumatin gene and the promoter sequence and glucoamylase signal sequence of Aspergillus niger. The trpC terminator sequence of Aspergillus nidulans was also present in this fragment.

The aforementioned 5.5 Kb fragment was ligated with a plasmid containing the sulfanilamide resistance sequence, previously cut with XbaI (the only cutting site on this plasmid). The ligating mix was used to transform E. coli, strain DH5αF'. The resulting plasmid was called pThIII, as indicated in Figure 9.

The pThIII plasmid contained: (i) the synthetic gene which codifies thaumatin II under the control of the glucoamylase promoter of Aspergillus niger; (ii) the signal sequence ("pre") and the "pro" sequence of the glucoamylase gene of Aspergillus niger; (iii) a sulfanilamide resistance marker; and (iv) the trpC terminator of Aspergillus nidulans. The final identity of this construction was verified by sequencing.

A strain of Penicillium roquefortii was transformed with plasmid pThIII according to the same method described in Example 1 (sections 1.3.1 and 1.3.2). The colonies resistant to sulfanilamide were tested to see if their genomes contained the substantially modified gene codifying thaumatin II. The methods used (PCR) were analogous to those described in Example 1 (section 1.4.1).

Figure 10 shows the result of a PCR experiment. The two oligonucleotides used to detect the thaumatin gene were the same ones used before (T1 and T2). With these two oligonucleotides, a fragment of 604 pairs of bases can be amplified corresponding to nucleotides 21 to 624 of the synthetic gene encoding thaumatin II. Figure 10 shows that when DNA from an untransformed fungus ("control", lane 2) is used, none of the bands corresponding to the synthetic gene are amplified, whereas when DNA is used from a fungus transformed with pThIII, a band of the expected size is amplified (lane 3). This fungus was called transformant a2. For control purposes, the reaction products obtained when plasmid pThIII was used were also run through the gel (lane 4).

The figure shows that transformant a2 correctly incorporated the synthetic gene of thaumatin II in its genome. Therefore, it was analyzed in greater detail to see if it expressed and secreted thaumatin II correctly. For immunoblotting analysis (Western-Blot) of the recombinant thaumatin, the methods described in section (1.4.2.) were used with the following modifications.

The experiment was started with 1 liter of a2 strain of Penicillium roquefortii which was grown for 8 days at 28°C in a semidefined medium for mycelium production (MSDPM). After vacuum filtration with a Büchner funnel, producing 45 g of mycelium per liter of culture, both the culture medium (liquid fraction) and the retained mycelium (solid fraction, 4.5 g) were analyzed.

The solid fraction was processed using the methods outlined in section (1.4.2.1), including sonication, thus obtaining 13.5 mL of mycelium extract in sonication solution.

The 13.5 mL of mycelium extract and 10 mL of culture medium were precipitated with 10% trichloracetic acid and the precipitated material was resuspended in a final volume of 200 µL of sonication solution. These samples were then analyzed by protein electrophoreses and immunoblotting as described in detail in Example 1, section (1.4.2).

The results of this experiment are shown in Figure 11 (14% SDS-polyacrylamide gel). Lane 7 in this figure contains proteins of known molecular weight (markers). The molecular weight corresponding to each protein is listed on the right of the figure. Lane 2 contains a sample of culture medium where fungus T0901 was grown. As described in Example 1, this fungus is a producer of intracellular thaumatin. Lanes 3 and 5 contain samples of culture medium (E for extracellular) and mycelium (I for intracellular) corresponding to transformant a2. Lanes 4 and 6 contain the same samples (E and I) corresponding to untransformed Penicillium roquefortii. As is clearly seen in Figure 11, transformant a2 turned out to be a good producer and secretor of thaumatin.

However, the effectiveness of the secretion was not complete given that a part of the thaumatin produced was not secreted, as is seen in the comparison between lanes 3 and 5. Organoleptic tests were performed on the culture broth and the characteristic sweet taste of thaumatin was detected.

EXAMPLE 3: EXTRACELLULAR PRODUCTION OF THAUMATIN IN THE AWAMORI VARIANT OF ASPERGILLUS NIGER

Strain NRRL312 of the awamori variant of Aspergillus niger was transformed in the presence of polyethylene glycol, as described in the literature (Yelton et al., Proc. Natl. Acad. Sci. USA, 1984, vol. 81, pp. 1470-4), with some modifications.

Four hundred mL of CM medium (malt extract, 5 g/L; yeast extract, 5 g/L; glucose, 5 g/L) in a 2-liter flask was inoculated with spores of the awamori variant of Aspergillus niger from a dish. The fungus grew for 16 hours. The mycelium was collected by filtration through a sterile gauze and washed with 100 mL of wash buffer (0.6 M $MgSO_4$, 10 mM Na_3PO_4, pH 5.8). The mycelium was pressed in sterile paper towels and produced 2.5 grams.

For the formation of protoplasts, the mycelium was resuspended in 15 mL/g of cold protoplast buffer (1.2 M $MgSO_4$, 10 mM Na_3PO_4, pH 5.8). At this point, 40 mg of Lysin enzyme (Sigma) was added per g of mycelium and the mixture was placed in ice for five minutes. After this incubation, 1 mL of BSA solution was added (12 mg/mL in protoplast buffer) and the solution was incubated for 3 or 4 hours at 30°C. Protoplast formation was monitored using a microscope. The mixture was filtered through nylon or a glass membrane and washed with the protoplast buffer. The protoplasts were centrifuged at 2000 rpm at 4°C for 15 minutes with a floating rotor (Beckman GPR centrifuge). The protoplasts were resuspended in 15 mL of ST solution (1M sorbitol, 10 mN Tris-HCl, pH 7.5), centrifuged again and resuspended in 1 mL of ST. The solution was centrifuged again and washed twice with 1 mL of STC (ST plus 0.01 M $CaCl_2$). The protoplasts were counted under the microscope, centrifuged again and resuspended in sufficient volume of STC to obtain a concentration of 10^8 protoplasts/mL. Each 400-mL culture generally produced 10^8 protoplasts. At that point, the protoplasts were directly plated in regeneration medium, in 5-mL tubes of 0.7% soft agar with saccharose osmotic stabilizer (1M), and were plated in basal medium with 1.5% agar.

For the transformation experiments, 200 µL of the 10^8-protoplasts/mL protoplast solution was used to start. Ten µg of transformant DNA (pThIII in this case) and 50 µL of PTC (60% PEG 6000; 10 mN Tris-HCl, pH 7.5; 10 mM $CaCl_2$) were added to the protoplasts and the solution was incubated in ice for 20 minutes. One mL of PTC was then added and the solution was mixed well and kept at room temperature for five minutes. The protoplasts were centrifuged and resuspended in 200 µL of STC medium. The mixture was plated in regeneration medium with sulfanilamide at 1 mg/mL. The dishes were incubated upside down at 30°C. Regeneration was observed after three or four days of incubation.

(3.1) Preparation of the regeneration medium

1. Trace solution: 400 mg/L $CuSO_4 \cdot 5H_2O$; 800 mg/L $FeSO_4 \cdot 7H_2O$; 800 mg/L $MnSO_4 \cdot 2H_2O$; 800 mg/L $Na_2MoO_4 \cdot 2H_2O$; 40 mg/L $Na_2BrO_7 \cdot 10H_2O$; 8 mg/L $ZnSO_4 \cdot 7H_2O$.

2. Salt solution (50X): 26 g/L KCl; 26 g/L $MgSO_4 \cdot 7H_2O$; 76 g/L KH_2PO_4; 50 mL/L of trace solution.

3. Ammonium tartrate: 30 grams per liter.

4. MMA (minimum Aspergillus medium): 10 or 15 g of glucose, or 7 g of agar was added to 970 mL of distilled water (final concentrations of 1.5% or 0.7%, respectively). The mixture was autoclaved and 10 mL of sterile ammonium tartrate solution and 20 mL of sterile salt solution were then added. Finally, the regeneration medium was prepared by adding saccharose to the MMA medium until the concentration of 1 M was reached.

EXAMPLE IV: PRODUCTION, SECRETION AND PROCESSING OF A GLUCOAMYLASE-THAUMATIN FUSION PROTEIN

As outlined in Figure 12, the pGEX-KG plasmid (5.0 Kb) (Pharmacia Biotech) was sequentially controlled with NcoI and Hind III, thus generating a fragment of approximately 4900 bp. This fragment, which no longer contained the SalI restriction site of the pGEX-KG polylinker, was purified in a 0.8% agarose gel.

The previous fragment was ligated with a NcoI-HindIII fragment from plasmid pTZ18RN(3/4) which contained the second part of the synthetic gene of thaumatin, thus generating plasmid pECThI of approximately 5.3 Kb. This new plasmid was treated with NcoI and the linearized fragment was ligated with a NcoI-NcoI fragment from plasmid pTZ18RN(1/2), which contained the first part of the synthetic gene of thaumatin, thus generating plasmid pECThII (of approximately 5.6 Kb). Plasmid pECThII contained the synthetic gene of thaumatin under the control of the tac promoter of Escherichia coli. This construction made it possible to obtain intracellular production of recombinant thaumatin in Escherichia coli.

The starting point for the construction of pThIX was the pECThI plasmid (of approximately 5.3 Kb). To eliminate the only MscI restriction site present in this plasmid, pECThI was sequentially treated with MscI and EcoRV (enzymes which produce blunt ends), thus releasing two fragments of 4.1 Kb and 1.2 Kb. The 4.1-Kb fragment was purified in a 0.8% agarose gel and religated through the action of DNA ligase. The result was plasmid pThIV. This plasmid was linearized with NcoI and the linear fragment was ligated with a NcoI-NcoI fragment from plasmid pTZ18RN(1/2), which contained the first half of the synthetic gene of thaumatin, thus generating plasmid pThV.

The single-stranded oligonucleotides, GLA1 and GLA2, were commercially bought (Ingenasa S.A) and have the following sequences (included in those of Figure 14):

GLA1: 5'-AATTCTGCGGAACGTCGACCGCGACGGTGACTGACACCTGGCGGC GAATGGATAAAAGGG-3'

GLA2: 5'-CCCTTTTATCCATTCGCCGCCAGGTGTCAGTCACCGTCGCGGTCG ACGTTCCGCAG-3'

These two oligonucleotides were annealed as follows: 10µg of each oligonucleotide was mixed in ligation buffer (40 mM Tris-HCl, pH 7.5; 20 mM $MgCl_2$; 50 mM NaCl) in a final volume of 25 µL. The mixture was heated for 5 minutes at 65°C and the temperature was allowed to drop slowly (for one half hour) to 30°C. The double-stranded DNA annealed in this way was purified in a 8% polyacrilamide gel. This double-stranded synthetic oligonucleotide, called GLA(1/2), had one blunt edge and one EcoRI end. Plasmid pThV was digested with MscI and EcoRI and ligated with the GLA(1/2) synthetic fragment, thus generating pThVI. Figure 14 shows the connection between the last sequences of the glucoamylase gene of Aspergillus niger, the spacer sequence and the synthetic gene of thaumatin II.

The next step was to insert the complete gene of glucoamylase (glaA) of Aspergillus niger or the awamori variant of Aspergillus niger, respectively, in phase with the complete gene of thaumatin II so that a glucoamylase-thaumatin fusion protein could be formed.

Plasmid pFGA2, obtained from the Belgian collection of cultures and LMBP plasmids (Ghent, Belgium, number 1728), contained the complete gene of glucoamylase (glaA) of Aspergillus niger. The plasmid was digested with EcoRI and SalI, and a fragment of approximately 2.3 Kb was isolated containing the complete gene of glucoamylase except for the last 10 amino acids of the protein. This fragment was ligated with plasmid pThVI which had previously been digested with EcoRI and SalI, thus generating plasmid pThVII (the junctions are described in Figure 14).

To obtain the glucoamylase gene of the awamori variant of Aspergillus niger, the following process was followed: total DNA of the NRRL312 strain of this fungus was prepared according to the protocol in section (1.4.1.1) . Two oligonucleotides, complementary to the 5' and 3' ends of the glucoamylase gene were used to amplify the complete gene. The fragment thus amplified was purified in a 0.8% agarose gel and digested with EcoRI and SalI. This 2.3-Kb EcoRI-SalI fragment was subcloned in pBluescript SK (Stratagene Inc.), which had previously been digested with EcoRI and SalI, thus generating plasmid pGLA-Aw.

In order to place the glucoamylase-spacer-thaumatin cassette under the control of the gla promoter of Aspergillus niger, the pThVII plasmid was digested with the restriction enzymes BssHII (partial digestion) and HindIII, and a fragment of approximately 3.0 Kb was isolated. This fragment was ligated with pAN52-6B which had previously been digested with BssHII and HindIII, thus obtaining plasmid pThVIII. Finally, the sulfanilamide resistance gene (Su^R) was inserted as described in Example 2, thus generating pThIX.

Plasmid pThIX contained: (i) a sulfanilamide resistance marker; (ii) a DNA sequence which encodes a fusion protein formed by (a) the synthetic gene of thaumatin II, (b) a spacer sequence which in turn contains a KEX2 processing sequence, and (c) the complete glucoamylase gene of Aspergillus niger; and finally, (iii) the signal sequence ("pre") and the "pro" sequence of the glucoamylase gene (glaA) of Aspergillus niger.

Plasmid pThIX was used to transform the awamori variant of Aspergillus niger as per the protocols specified in Example 3. Transformants which correctly secreted and processed thaumatin were obtained, and it was determined that the protein was sweet.

In the same way, but using plasmid pGLA-Aw instead of plasmid pThVII, an analogue plasmid of pThIX was obtained containing the gla sequence of A. awamori instead of that of A. niger. Similarly, this plasmid was also used to transform a strain of A. awamori, with similar results.

LIST OF SEQUENCES

SEQ. ID No.1

GCC	ACC	TTC	GAG	ATC	GTC	AAC	CGC	TGC	TCC	TAC	ACC	GTC	TGG	GCG	GCC	48
Ala	Thr	Phe	Glu	Ile	Val	Asn	Arg	Cys	Ser	Tyr	Thr	Val	Trp	Ala	Ala	
1				5					10					15		
GCC	TCC	AAA	GGC	GAC	GCC	GCC	CTG	GAC	GCC	GGC	GGC	CGC	CAG	CTC	AAC	96
Ala	Ser	Lys	Gly	Asp	Ala	Ala	Leu	Asp	Ala	Gly	Gly	Arg	Gln	Leu	Asn	
			20					25					30			
TCG	GGA	GAG	TCC	TGG	ACC	ATC	AAC	GTA	GAA	CCC	GGC	ACC	AAG	GGC	GGC	144
Ser	Gly	Glu	Ser	Trp	Thr	Ile	Asn	Val	Glu	Pro	Gly	Thr	Lys	Gly	Gly	
		35					40					45				
AAA	ATC	TGG	GCC	CGC	ACC	GAC	TGC	TAT	TTC	GAC	GAC	AGC	GGC	CGC	GGC	192
Lys	Ile	Trp	Ala	Arg	Thr	Asp	Cys	Tyr	Phe	Asp	Asp	Ser	Gly	Arg	Gly	
	50					55					60					
ATC	TGC	CGG	ACC	GGC	GAC	TGC	GGC	GGC	CTC	CTC	CAG	TGC	AAG	CGC	TTC	240
Ile	Cys	Arg	Thr	Gly	Asp	Cys	Gly	Gly	Leu	Leu	Gln	Cys	Lys	Arg	Phe	
65					70					75					80	
GGC	CGG	CCG	CCC	ACC	ACG	CTG	GCG	GAC	TTC	TCG	CTC	AAC	CAG	TAC	GGC	288
Gly	Arg	Pro	Pro	Thr	Thr	Leu	Ala	Glu	Phe	Ser	Leu	Asn	Gln	Tyr	Gly	
				85					90					95		
AAG	GAC	TAC	ATC	GAC	ATC	TCC	AAC	ATC	AAA	GGC	TTC	AAC	GTG	CCG	ATC	336
Lys	Asp	Tyr	Ile	Asp	Ile	Ser	Asn	Ile	Lys	Gly	Phe	Asn	Val	Pro	Met	
			100					105					110			
GAC	TTC	TGC	CCG	ACC	ACG	CGC	GGC	TGC	CGC	GGG	GTG	CGG	TGC	GCC	GCC	384
Asp	Phe	Ser	Pro	Thr	Thr	Arg	Gly	Cys	Arg	Gly	Val	Arg	Cys	Ala	Ala	
		115					120					125				
GAC	ATC	GTG	GGC	CAG	TGC	CCG	GCG	AAG	CTG	AAG	GCG	CCG	GGG	GGT	GGT	432
Asp	Ile	Val	Gly	Gln	Cys	Pro	Ala	Lys	Leu	Lys	Ala	Pro	Gly	Gly	Gly	
	130					135					140					
TGC	AAC	GAT	GCC	TGC	ACC	GTG	TTC	CAG	ACG	AGC	GAG	TAC	TGC	TGC	ACC	480
Cys	Asn	Asp	Ala	Cys	Thr	Val	Phe	Gln	Thr	Ser	Glu	Tyr	Cys	Cys	Thr	
145					150					155					160	
ACG	GGG	AAG	TGC	GGG	CCG	ACG	GAG	TAC	TCG	CGC	TTC	TTC	AAG	AGG	CTT	528
Thr	Gly	Lys	Cys	Gly	Pro	Thr	Glu	Tyr	Ser	Arg	Phe	Phe	Lys	Arg	Leu	
				165					170					175		
TGC	CCC	GAC	GCC	TTC	AGT	TAT	GTC	CTG	GAC	AAG	CCA	ACC	ACC	GTC	ACC	576
Cys	Pro	Asp	Ala	Phe	Ser	Tyr	Val	Leu	Asp	Lys	Pro	Thr	Thr	Val	Thr	
			180					185					190			
TGC	CCC	GGC	AGC	TCC	AAC	TAC	AGC	GTC	ACT	TTC	TGC	CCT	ACT	GCC(TAA).		624
Cys	Pro	Gly	Ser	Ser	Asn	Tyr	Arg	Val	Thr	Phe	Cys	Pro	Thr	Ala		
		195					200					205				

Sequence ID No. 1: Amino-acid sequence of the protein thaumatin II, and nucleotide sequence of the natural gene.

SEQ. ID No.2

```
GCC ACC TTC GAG ATC GTC AAC CGC TGC TCC TAC ACC GTC TGG GCC GCC   48
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15

GCC TCC AAG GGC GAC GCC GCC CTC GAC GCC GGC GGC CGC CAG CTC AAC   96
Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

TCC GGC GAG TCC TGG ACC ATC AAC GTC GAG CCC GGC ACC AAG GGC GGC  144
Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Lys Gly Gly
        35                  40                  45

AAG ATC TGG GCC CGC ACC GAC TGC TAC TTC GAC GAC TCC GGC CGC GGC  192
Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Arg Gly
    50                  55                  60

ATC TGC CGC ACC GGC GAC TGC GGC GGC CTC CTC CAG TGC AAG CGC TTC  240
Ile Cys Arg Thr Gly Asp Cys Gly Gly Leu Leu Gln Cys Lys Arg Phe
65                  70                  75                  80

GGC CGC CCC CCC ACC ACC CTC GCC GAC TTC TCC CTC AAC CAG TAC GGC  288
Gly Arg Pro Pro Thr Thr Leu Ala Glu Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

AAG GAC TAC ATC GAC ATC TCC AAC ATC AAG GGC TTC AAC GTC CCC ATG  336
Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

GAC TTC TCC CCC ACC ACC CGC GGC TGC CGC GGC GTC CGC TGC GCC GCC  384
Asp Phe Ser Pro Thr Thr Arg Gly Cys Arg Gly Val Arg Cys Ala Ala
        115                 120                 125

GAC ATC GTC GGC CAG TGC CCC GCC AAG CTC AAG GCC CCC GGC GGC GGC  432
Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140

TGC AAC GAC GCC TGC ACC GTC TTC CAG ACC TCC GAG TAC TGC TGC ACC  480
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

ACC GGC AAG TGC GGC CCC ACC GAG TAC TCC CGC TTC TTC AAG CGC CTC  528
Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

TGC CCC GAC GCC TTC TCC TAC GTC CTC GAC AAG CCC ACC ACC GTC ACC  576
Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

TGC CCC GGC TCC TCC AAC TAC CGC GTC ACC TTC TGC CCC ACC GCC(TAA)n  624
Cys Pro Gly Ser Ser Asn Tyr Arg Val Thr Phe Cys Pro Thr Ala
        195                 200                 205
```

Sequence ID No. 2: Amino-acid sequence of thaumatin II and nucleotide sequence of the artificial, synthetic and completely optimized gene, used in the examples of this invention, to which the n codons with TAA termination ($n \geq 1$) were added.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: URQUIMA,S.A.
(B) STREET: Dega Bahi, 59-67
(C) CITY: Barcelona
(D) STATE: Barcelona
(E) COUNTRY: Spain
(F) POSTAL CODE (ZIP): 08026
(G) TELEPHONE: 343-3471511
(H) TELEFAX: 343-4560639
(I) TELEX: 52.963URIAC E

(ii) TITLE OF INVENTION: Preparation of a natural protein sweetener

(iii) NUMBER OF SEQUENCES: 22

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER: EP 95 105 973.2

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: ES 9400836
(B) FILING DATE: 21-APR-1994

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Thaumatoccus daniellii
(D) DEVELOPMENTAL STAGE: Adult
(F) TISSUE TYPE: Arils
(G) CELL TYPE: Pollen mother cell
(I) ORGANELLE: Cyanelle

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..621

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GCC ACC TTC GAG ATC GTC AAC CGC TGC TCC TAC ACC GTG TGG GCG GCC      48
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15

GCC TCC AAA GGC GAC GCC GCC CTG GAC GCC GGC GGC CGC CAG CTC AAC      96
Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

TCG GGA GAG TCC TGG ACC ATC AAC GTA GAA CCC GGC ACC AAG GGC GGC     144
Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Lys Gly Gly
        35                  40                  45

AAA ATC TGG GCC CGC ACC GAC TGC TAT TTC GAC GAC AGC GGC CGC GGC     192
Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Arg Gly
    50                  55                  60

ATC TGC CGG ACC GGC GAC TGC GGC GGC CTC CTC CAG TGC AAG CGC TTC     240
Ile Cys Arg Thr Gly Asp Cys Gly Gly Leu Leu Gln Cys Lys Arg Phe
65                  70                  75                  80

GGC CGG CCG CCC ACC ACG CTG GCG GAC TTC TCG CTC AAC CAG TAC GGC     288
Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

AAG GAC TAC ATC GAC ATC TCC AAC ATC AAA GGC TTC AAC GTG CCG ATG     336
Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

GAC TTC TGC CCG ACC ACG CGC GGC GGC TGC CGC GGG GTG CGG TGC GCC     384
Asp Phe Cys Pro Thr Thr Arg Gly Gly Cys Arg Gly Val Arg Cys Ala
        115                 120                 125

GAC ATC GTG GGC CAG TGC CCG GCG AAG CTG AAG GCG CCG GGG GGT GGT     432
Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140

TGC AAC GAT GCG TGC ACC GTG TTC CAG ACG AGC GAG TAC TGC TGC ACC     480
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

ACG GGG AAG TGC GGG CCG ACG GAG TAC TCG CGC TTC TTC AAG AGG CTT     528
Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

TGC CCC GAC GCG TTC AGT TAT GTC CTG GAC AAG CCA ACC ACC GTC ACC     576
Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

TGC CCC GGC AGC TCC AAC TAC AGC GTC ACT TTC TGC CCT ACT GCC         621
Cys Pro Gly Ser Ser Asn Tyr Ser Val Thr Phe Cys Pro Thr Ala
        195                 200                 205

TAA                                                                 624
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15

Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Lys Gly Gly
        35                  40                  45

Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Arg Gly
    50                  55                  60

Ile Cys Arg Thr Gly Asp Cys Gly Gly Leu Leu Gln Cys Lys Arg Phe
65                  70                  75                  80

Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

Asp Phe Cys Pro Thr Thr Arg Gly Gly Cys Arg Gly Val Arg Cys Ala
        115                 120                 125

Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140

Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

Cys Pro Gly Ser Ser Asn Tyr Ser Val Thr Phe Cys Pro Thr Ala
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "The molecule is a synthetic, double-stranded DNA sequence. It was constructed by the assembly of several oligonucleotides. Codon usage is designed for optimal expression in filamentous fungi."

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Thaumatoccocus danielli
(D) DEVELOPMENTAL STAGE: Adult
(F) TISSUE TYPE: Arils
(G) CELL TYPE: Pollen mother cell
(I) ORGANELLE: Cyanelle

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..621

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
GCC ACC TTC GAG ATC GTC AAC CGC TGC TCC TAC ACC GTC TGG GCC GCC          48
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15

GCC TCC AAG GGC GAC GCC GCC CTC GAC GCC GGC GGC CGC CAG CTC AAC          96
Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

TCC GGC GAG TCC TGG ACC ATC AAC GTC GAG CCC GGC ACC AAG GGC GGC         144
Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Lys Gly Gly
        35                  40                  45

AAG ATC TGG GCC CGC ACC GAC TGC TAC TTC GAC GAC TCC GGC CGC GGC         192
Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Arg Gly
    50                  55                  60

ATC TGC CGC ACC GGC GAC TGC GGC GGC CTC CTC CAG TGC AAG CGC TTC         240
Ile Cys Arg Thr Gly Asp Cys Gly Gly Leu Leu Gln Cys Lys Arg Phe
65                  70                  75                  80

GGC CGC CCC CCC ACC ACC CTC GCC GAC TTC TCC CTC AAC CAG TAC GGC         288
Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

AAG GAC TAC ATC GAC ATC TCC AAC ATC AAG GGC TTC AAC GTC CCC ATG         336
Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

GAC TTC TCC CCC ACC ACC CGC GGC TGC CGC GGC GTC CGC TGC GCC GCC         384
Asp Phe Ser Pro Thr Thr Arg Gly Cys Arg Gly Val Arg Cys Ala Ala
        115                 120                 125

GAC ATC GTC GGC CAG TGC CCC GCC AAG CTC AAG GCC CCC GGC GGC GGC         432
Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140

TGC AAC GAC GCC TGC ACC GTC TTC CAG ACC TCC GAG TAC TGC TGC ACC         480
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

ACC GGC AAG TGC GGC CCC ACC GAG TAC TCC CGC TTC TTC AAG CGC CTC         528
Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

TGC CCC GAC GCC TTC TCC TAC GTC CTC GAC AAG CCC ACC ACC GTC ACC         576
Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190
```

```
TGC CCC GGC TCC TCC AAC TAC CGC GTC ACC TTC TGC CCC ACC GCC          621
Cys Pro Gly Ser Ser Asn Tyr Arg Val Thr Phe Cys Pro Thr Ala
        195                 200                 205

TAA                                                                  624
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15

Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Lys Gly Gly
        35                  40                  45

Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Arg Gly
    50                  55                  60

Ile Cys Arg Thr Gly Asp Cys Gly Gly Leu Leu Gln Cys Lys Arg Phe
65                  70                  75                  80

Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

Asp Phe Ser Pro Thr Thr Arg Gly Cys Arg Gly Val Arg Cys Ala Ala
        115                 120                 125

Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140

Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

Cys Pro Gly Ser Ser Asn Tyr Arg Val Thr Phe Cys Pro Thr Ala
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Thaumatococcus daniellii
(D) DEVELOPMENTAL STAGE: Adult
(F) TISSUE TYPE: Arils
(G) CELL TYPE: Pollen mother cell
(I) ORGANELLE: Cyanelle

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..621

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
GCC ACC TTC GAG ATC GTC AAC CGC TGC TCC TAC ACC GTG TGG GCG GCC          48
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15

GCC TCC AAA GGC GAC GCC GCC CTG GAC GCC GGC GGC CGC CAG CTC AAC          96
Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

TCG GGA GAG TCC TGG ACC ATC AAC GTA GAA CCC GGC ACC AAC GGC GGC         144
Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Asn Gly Gly
        35                  40                  45

AAA ATC TGG GCC CGC ACC GAC TGC TAT TTC GAC GAC AGC GGC TCC GGC         192
Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Ser Gly
    50                  55                  60

ATC TGC AAG ACC GGC GAC TGC GGC GGC CTC CTC CGC TGC AAG CGC TTC         240
Ile Cys Lys Thr Gly Asp Cys Gly Gly Leu Leu Arg Cys Lys Arg Phe
65                  70                  75                  80

GGC CGG CCG CCC ACC ACG CTG GCG GAC TTC TCG CTC AAC CAG TAC GGC         288
Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

AAG GAC TAC ATC GAC ATC TCC AAC ATC AAA GGC TTC AAC GTG CCG ATG         336
Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

AAC TTC TGC CCG ACC ACG CGC GGC GGC TGC CGC GGG GTG CGG TGC GCC         384
Asn Phe Cys Pro Thr Thr Arg Gly Gly Cys Arg Gly Val Arg Cys Ala
        115                 120                 125
```

```
GAC ATC GTG GGC CAG TGC CCG GCG AAG CTG AAG GCG CCG GGG GGT GGT          432
Asp Ile Val Gly Gln Cys Pro Ala Lys Leu L  Ala Pro Gly Gly Gly
    130                 135                 140

TGC AAC GAT GCG TGC ACC GTG TTC CAG ACG AGC GAG TAC TGC TGC ACC          480
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

ACG GGG AAG TGC GGG CCG ACG GAG TAC TCG CGC TTC TTC AAG AGG CTT          528
Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

TGC CCC GAC GCG TTC AGT TAT GTC CTG GAC AAG CCA ACC ACC GTC ACC          576
Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

TGC CCC GGC AGC TCC AAC TAC AGC GTC ACT TTC TGC CCT ACT GCC              621
Cys Pro Gly Ser Ser Asn Tyr Ser Val Thr Phe Cys Pro Thr Ala
        195                 200                 205

TAA                                                                      624
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
  1               5                  10                  15

Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Asn Gly Gly
        35                  40                  45

Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Ser Gly
    50                  55                  60

Ile Cys Lys Thr Gly Asp Cys Gly Gly Leu Leu Arg Cys Lys Arg Phe
65                  70                  75                  80

Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

Asn Phe Cys Pro Thr Thr Arg Gly Gly Cys Arg Gly Val Arg Cys Ala
        115                 120                 125

Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140
```

```
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

Cys Pro Gly Ser Ser Asn Tyr Ser Val Thr Phe Cys Pro Thr Ala
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "The molecule is a synthetic, double-stranded DNA sequence. It was constructed by the assembly of several oligonucleotides. Codon usage is designed for optimal expression in filamentous fungi."

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Thaumatococcus daniellii
(D) DEVELOPMENTAL STAGE: Adult
(F) TISSUE TYPE: Arils
(G) CELL TYPE: Pollen mother cell
(I) ORGANELLE: Cyanelle

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..621

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
GCC ACC TTC GAG ATC GTC AAC CGC TGC TCC TAC ACC GTC TGG GCC GCC        48
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
  1               5                  10                  15

GCC TCC AAG GGC GAC GCC GCC CTC GAC GCC GGC GGC CGC CAG CTC AAC        96
Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

TCC GGC GAG TCC TGG ACC ATC AAC GTC GAG CCC GGC ACC AAC GGC GGC       144
Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Asn Gly Gly
        35                  40                  45
```

```
AAG ATC TGG GCC CGC ACC GAC TGC TAC TTC GAC GAC TCC GGC TCC GGC        192
Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Ser Gly
     50                  55                  60

ATC TGC AAG ACC GGC GAC TGC GGC GGC CTC CTC CGC TGC AAG CGC TTC        240
Ile Cys Lys Thr Gly Asp Cys Gly Gly Leu Leu Arg Cys Lys Arg Phe
 65                  70                  75                  80

GGC CGC CCC CCC ACC ACC CTC GCC GAC TTC TCC CTC AAC CAG TAC GGC        288
Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

AAG GAC TAC ATC GAC ATC TCC AAC ATC AAG GGC TTC AAC GTC CCC ATG        336
Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

AAC TTC TCC CCC ACC ACC CGC GGC TGC CGC GGC GTC CGC TGC GCC GCC        384
Asn Phe Ser Pro Thr Thr Arg Gly Cys Arg Gly Val Arg Cys Ala Ala
        115                 120                 125

GAC ATC GTC GGC CAG TGC CCC GCC AAG CTC AAG GCC CCC GGC GGC GGC        432
Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140

TGC AAC GAC GCC TGC ACC GTC TTC CAG ACC TCC GAG TAC TGC TGC ACC        480
Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

ACC GGC AAG TGC GGC CCC ACC GAG TAC TCC CGC TTC TTC AAG CGC CTC        528
Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

TGC CCC GAC GCC TTC TCC TAC GTC CTC GAC AAG CCC ACC ACC GTC ACC        576
Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

TGC CCC GGC TCC TCC AAC TAC CGC GTC ACC TTC TGC CCC ACC GCC            621
Cys Pro Gly Ser Ser Asn Tyr Arg Val Thr Phe Cys Pro Thr Ala
        195                 200                 205

TAA                                                                    624
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Ala Thr Phe Glu Ile Val Asn Arg Cys Ser Tyr Thr Val Trp Ala Ala
 1               5                  10                  15

Ala Ser Lys Gly Asp Ala Ala Leu Asp Ala Gly Gly Arg Gln Leu Asn
            20                  25                  30

Ser Gly Glu Ser Trp Thr Ile Asn Val Glu Pro Gly Thr Asn Gly Gly
        35                  40                  45

Lys Ile Trp Ala Arg Thr Asp Cys Tyr Phe Asp Asp Ser Gly Ser Gly
    50                  55                  60
```

```
Ile Cys Lys Thr Gly Asp Cys Gly Gly Leu Leu Arg Cys Lys Arg Phe
 65                  70                  75                  80

Gly Arg Pro Pro Thr Thr Leu Ala Asp Phe Ser Leu Asn Gln Tyr Gly
                85                  90                  95

Lys Asp Tyr Ile Asp Ile Ser Asn Ile Lys Gly Phe Asn Val Pro Met
            100                 105                 110

Asn Phe Ser Pro Thr Thr Arg Gly Cys Arg Gly Val Arg Cys Ala Ala
        115                 120                 125

Asp Ile Val Gly Gln Cys Pro Ala Lys Leu Lys Ala Pro Gly Gly Gly
    130                 135                 140

Cys Asn Asp Ala Cys Thr Val Phe Gln Thr Ser Glu Tyr Cys Cys Thr
145                 150                 155                 160

Thr Gly Lys Cys Gly Pro Thr Glu Tyr Ser Arg Phe Phe Lys Arg Leu
                165                 170                 175

Cys Pro Asp Ala Phe Ser Tyr Val Leu Asp Lys Pro Thr Thr Val Thr
            180                 185                 190

Cys Pro Gly Ser Ser Asn Tyr Arg Val Thr Phe Cys Pro Thr Ala
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: P115

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ACCCGGGGAT CCTCTCCATG GGACCTGCAG GCATGCA 37

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Corresponds to those nucleotides in pTZ18R (Pharmacia) that have been modified as per the patent application."

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: pTZ18RN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCCGGGGATC CTCTCCATGG GACCTGCAGG CAT 33

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 106 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 1a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AAATGGAGGA TCCATGGCCA CCTTCGAGAT CGTCAACCGC TGCTCCTACA CCGTCTGGGC 60

CGCCGCCTCC AAGGGCGACG CCGCCCTCGA CGCCGGCGGC CGCCAG 106

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 1b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GCGGGCCCAG ATCTTGCCGC CCTTGGTGCC GGGCTCGACG TTGATGGTCC AGGACTCGCC 60

GGAGTTGAGC TGGCGGCCGC CGGCGTC 87

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 2a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GGCGGCAAGA TCTGGGCCCG CACCGACTGC TACTTCGACG ACTCCGGCCG CGGCATCTGC 60

CGCACCGGCG ACTGCGGCGG CCTCCTCCAG TGCAAGCGCT TCGGCCGCCC CCCCACC 117

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 2b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

AGTCCATGGG GACGTTGAAG CCCTTGATGT TGGAGATGTC GATGTAGTCC TTGCCGTACT 60

GGTTGAGGGA GAACTCGGCG AGGGTGGTGG GGGGGCGGCC GAA 103

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 84 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 3a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

AACGTCCCCA TGGACTTCTC CCCCACCACC CGCGGCTGCC GCGGCGTCCG CTGCGCCGCC 60

GACATCGTCG GCCAGTGCCC CGCC 84

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 3b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGACGGTGCA GGCGTCGTTG CAGCCGCCGC CGGGGGCCTT GAGCTTGGCG GGGCACTGGC 60

CGAC 64

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 4a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TCCAGACCTC CGAGTACTGC TGCACCACCG GCAAGTGCGG CCCCACCGAG TACTCCCGCT 60

TCTTCAAGCG CCTCTGCCCC GACGCCTTCT CCTACGTCCT C 101

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: 4b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GCTTGCCTGC AGTTATTATT AGGCGGTGGG GCAGAAGGTG ACGCGGTAGT TGGAGGAGCC 60

GGGGCAGGTG ACGGTGGTGG GCTTGTCGAG GACGTAGGAG AAGGCGT 107

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: T1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CCGCTGCTCC TACACCGTCT GGGCCG 26

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: T2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TTAGGCGGTG GGGCAGAAGG 20

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: GLA1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

AATTCTGCGG AACGTCGACC GCGACGGTGA CTGACACCTG GCGGCGAATG GATAAAAGGG 60

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: GLA2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CCCTTTTATC CATTCGCCGC CAGGTGTCAG TCACCGTCGC GGTCGACGTT CCGCAG 56

Claims

1. A DNA sequence which codifies the amino-acid sequence corresponding to the 207 amino acids of the protein thaumatin II (included in Sequence ID No. 1), followed by n stop sequences, where integer n is greater than or equal to 1; said DNA sequence being characterized in that it is the result of modifying, more than 50% of the possible, the DNA sequence of the natural gene which codifies the 207 amino acids of thaumatin II (natural gene also shown in Sequence ID No. 1); said modification consisting of adding n stop codons, where integer n is greater than or equal to one, and effecting more than 50% of the possible changes in the nucleotide codons corresponding to the amino acids of thaumatin II; said changes consisting of replacing the original codons in all the amino acids possible, with the codons indicated in parentheses in the following list of amino-acid codons:
Ala (GCC), Arg (CGC), Asn (AAC), Asp (GAC), Cys (TGC), Lys (AAG), Gln (CAG), Glu (GAG), Gly (GGC), Ile (ATC), Leu (CTC), Met (ATG), Phe (TTC), Pro (CCC), Ser (TCC), Thr (ACC), Trp (TGG), Tyr (TAC), Val (GTC).

2. A DNA sequence according to claim 1 where the modification consists of adding from one to three stop codons (n = 1, 2 or 3), and effecting more than 75% of the possible codon changes.

3. A DNA sequence according to claim 2 where all (100%) of the possible codon changes have been made so that the DNA sequence is the one in Sequence ID No. 2.

4. A DNA sequence according to any of claims 1, 2 or 3, wherein the stop codon(s) represent TAA.

5. A recombinant plasmid comprising: (i) a DNA sequence according to any of the claims 1 to 4; (ii) an expression cassette for filamentous fungi containing one promoter sequence and one terminating sequence for this type of fungi; (iii) an appropriate selection marker; and, optionally, (iv) a secretion signal DNA sequence for the extracellular production of the protein.

6. A recombinant plasmid according to claim 5 where the promoter sequence of the expression cassette comes from the gene of the enzyme glyceraldehyde 3-phosphate dehydrogenase of Aspergillus niger, or from the glucoamylase

gene of the same fungus; the terminating sequence of the expression cassette is that of tryptophan C of Aspergillus nidulans; and the selection marker is the sulfanilamide resistance marker.

7. A recombinant plasmid expressing the fusion protein thaumatin-glucoamylase comprising: (i) an appropriate selection marker; (ii) a DNA sequence made up of (a) a DNA sequence according to any of the claims 1 to 4, (b) a spacer sequence which in turn contains a KEX2 processing sequence, and (c) the complete glucoamylase gene of Aspergillus niger or the awamori variant of Aspergillus niger (glaA); and (iii) the "pre" signal sequence and the "pro" sequence of the glaA gene.

8. A filamentous fungus culture capable of producing the protein thaumatin II, which has been transformed with any of the plasmids in claims 5 to 7.

9. A culture according to claim 8 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

10. A process for producing thaumatin II comprising the following steps:

a) insertion of the DNA sequence from claims 1, 2, 3 or 4 in any of the expression vectors in claims 5, 6 and 7, using standard recombinant DNA technology techniques;

b) transformation of a strain of filamentous fungus with this expression vector;

c) culture of the strain of filamentous fungus which has been transformed in this way under the appropriate nutrient conditions, thus producing thaumatin II, either intracellularly, extracellularly or in both ways simultaneously, or in the form of the fusion protein thaumatin-glucoamylase.

d) depending on the case, separation and purification of thaumatin II alone, or separation of thaumatin II from the culture medium together with the fungus mycelium.

11. A process according to claim 10 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

12. A DNA sequence which codifies the amino-acid sequence corresponding to the 207 amino acids of the protein thaumatin I (207 amino acids which differ from those of Sequence ID No. 1 in only five amino acids, namely, 46-Asn, 63-Ser, 67-Lys, 76-Arg and 113-Asn), characterized in that it has the following five fixed codons: AAC (46-Asn), TCC (63-Ser), AAG (67-Lys), CGC (76-Arg) and AAC (113-Asn), and the rest of the codons are as in the DNA sequence in claim 1.

13. A DNA sequence according to claim 12 which has from one to three stop codons (n = 1, 2 or 3), and the rest of the codons which differ from the five fixed ones, as in the DNA sequence in claim 2.

14. A DNA sequence according to claim 13 which has the codons which are different from the five fixed ones, as in the DNA sequence in claim 3.

15. A recombinant plasmid comprising: (i) a DNA sequence according to any of the claims 12, 13 or 14; (ii) an expression cassette for filamentous fungi containing a promoter sequence and a terminating sequence which are appropriate for this type of fungi; (iii) an appropriate selection marker; and, optionally, (iv) a secretion signal DNA sequence for the extracelluar production of the protein.

16. A recombinant plasmid according to claim 15 where the promoter sequence of the expression cassette comes from the gene of the enzyme glyceraldehyde 3-phosphate dehydrogenase of Aspergillus niger, or from the glucoamylase gene of the same fungus; the terminating sequence of the expression cassette is tryptophane C of Aspergillus nidulans; and the selection marker is the sulfanilamide resistance selection marker.

17. A recombinant plasmid which expresses the fusion protein thaumatin-glucoamylase comprising: (i) an appropriate selection marker; (ii) a DNA sequence made up of (a) a DNA sequence according to claims 12, 13 or 14, (b) a spacer sequence which in turn contains a KEX2 processing sequence, and (c) the complete gene of glucoamylase of Aspergillus niger or the awamori variant of Aspergillus niger (glaA); and (iii) the "pre" signal sequence and the "pro" sequence from the glaA gene.

18. A filamentous fungus culture capable of producing the protein thaumatin I, which has been transformed with any of the plasmids in claims 15 to 17.

19. A culture according to claim 18 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

20. A process for producing thaumatin I comprising the following steps:

a) insertion of the DNA sequence of any of the claims 12, 13 or 14 in one expression vector selected from those in claims 15, 16 and 17, using standard recombinant DNA technology techniques;

b) transformation of a strain of filamentous fungus with this expression vector;

c) culture of the strain of filamentous fungus which has been transformed in this way under the appropriate nutrient conditions, thus producing thaumatin I either intracellularly, extracellularly or in both ways simultaneously, or in the form of the thaumatin-glucoamylase fusion protein.

d) depending on the case, separation and purification of thaumatin I alone, or separation of thaumatin I from the culture medium together with the fungus mycelium.

21. A process according to claim 20 where the filamentous fungus is selected from the species Penicillium roquefortii, Aspergillus niger, and the awamori variant of Aspergillus niger.

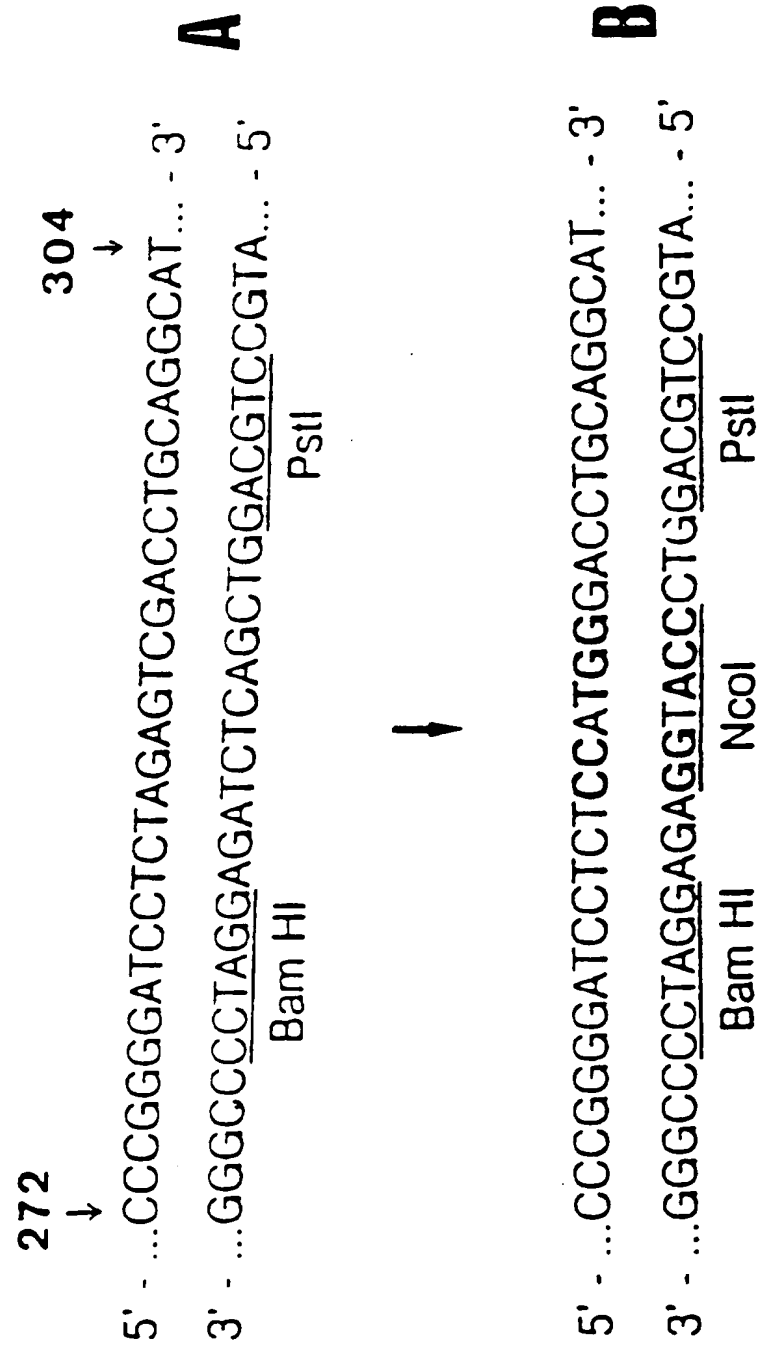


Fig. 1

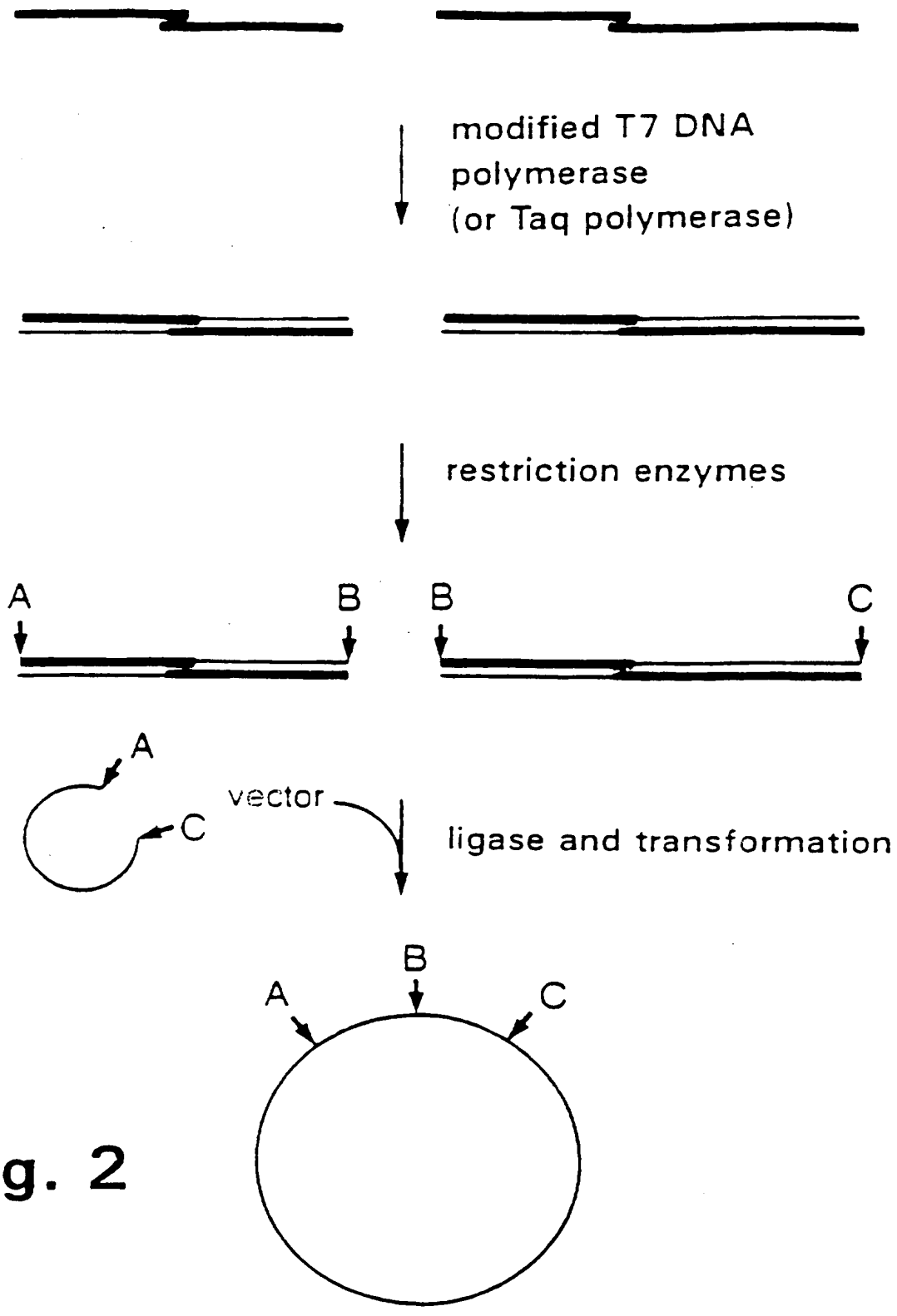


Fig. 2

1a (106 nucleotides).

5'- AAATGGAGGATCCATGGCCACCTTCGAGATCGTCAACCGC
TGCTCCTACACCGTCTGGGCCGCCGCCTCCAAGGGCGACGC
CGCCCTCGACGCCGGCGGCCGCCAG - 3'

1b (87 nucleotides).

5'- GCGGGCCCAGATCTTGCCGCCCTTGGTGCCGGGCTCGAC
GTTGATGGTCCAGGACTCGCCGGAGTTGAGCTGGCGGCCGC
CGGCGTC - 3'

2a (117 nucleotides).

5'- GGCGGCAAGATCTGGGCCCGCACCGACTGCTACTTCGACG
ACTCCGGCCGCGGCATCTGCCGCACCGGCGACTGCGGCGGC
CTCCTCCAGTGCAAGCGCTTCGGCCGCCCCCCCACC - 3'

2b (103 nucleotides).

5'- AGTCCATGGGGACGTTGAAGCCCTTGATGTTGGAGATGTCG
ATGTAGTCCTTGCCGTACTGGTTGAGGGAGAACTCGGCGAGGG
TGGTGGGGGGGCGGCCGAA - 3'

3a (84 nucleotides).

5'- AACGTCCCCATGGACTTCTCCCCCACCACCCGCGGCTGCCGCGGC
GTCCGCTGCGCCGCCGACATCGTCGGCCAGTGCCCCGCC - 3'

3b (64 nucleotides).

5'- AGACGGTGCAGGCGTCGTTGCAGCCGCCGCCGGGGGCCTTGAGCT
TGGCGGGGCACTGGCCGAC - 3'

4a (101 nucleotides).

5'- TCCAGACCTCCGAGTACTGCTGCACCACCGGCAAGTGCGGCCCCA
CCGAGTACTCCCGCTTCTTCAAGCGCCTCTGCCCCGACGCCTTCTCCT
ACGTCCTC - 3'

4b (107 nucleotides).

5'- GCTTGCCTGCAGTTATTATTAGGCGGTGGGGCAGAAGGTGACGCGG
TAGTTGGAGGAGCCGGGGCAGGTGACGGTGGTGGGCTTGTCGAGGAC
GTAGGAGAAGGCGT - 3'

Fig. 3

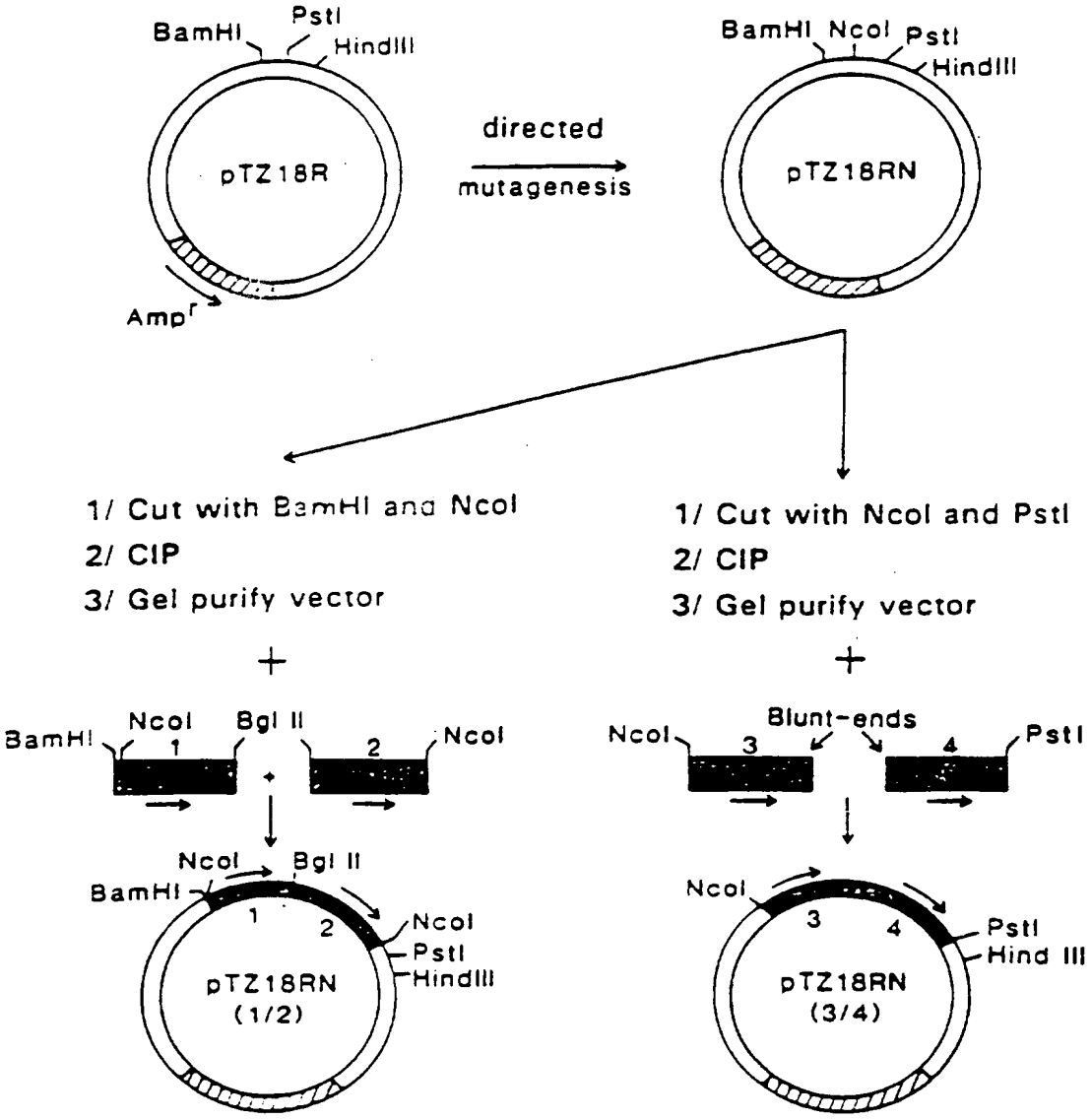


Fig. 4

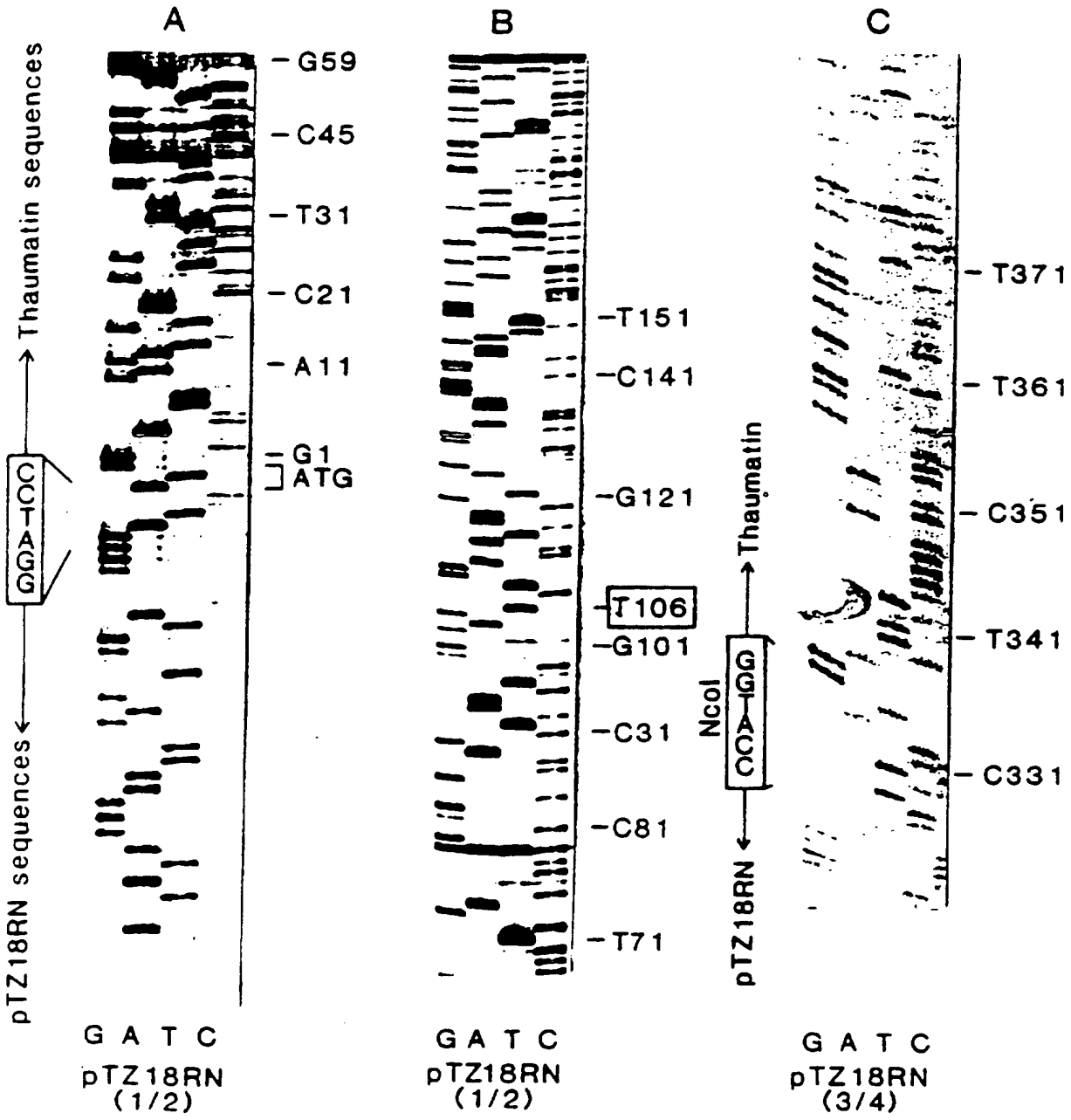


Fig. 5

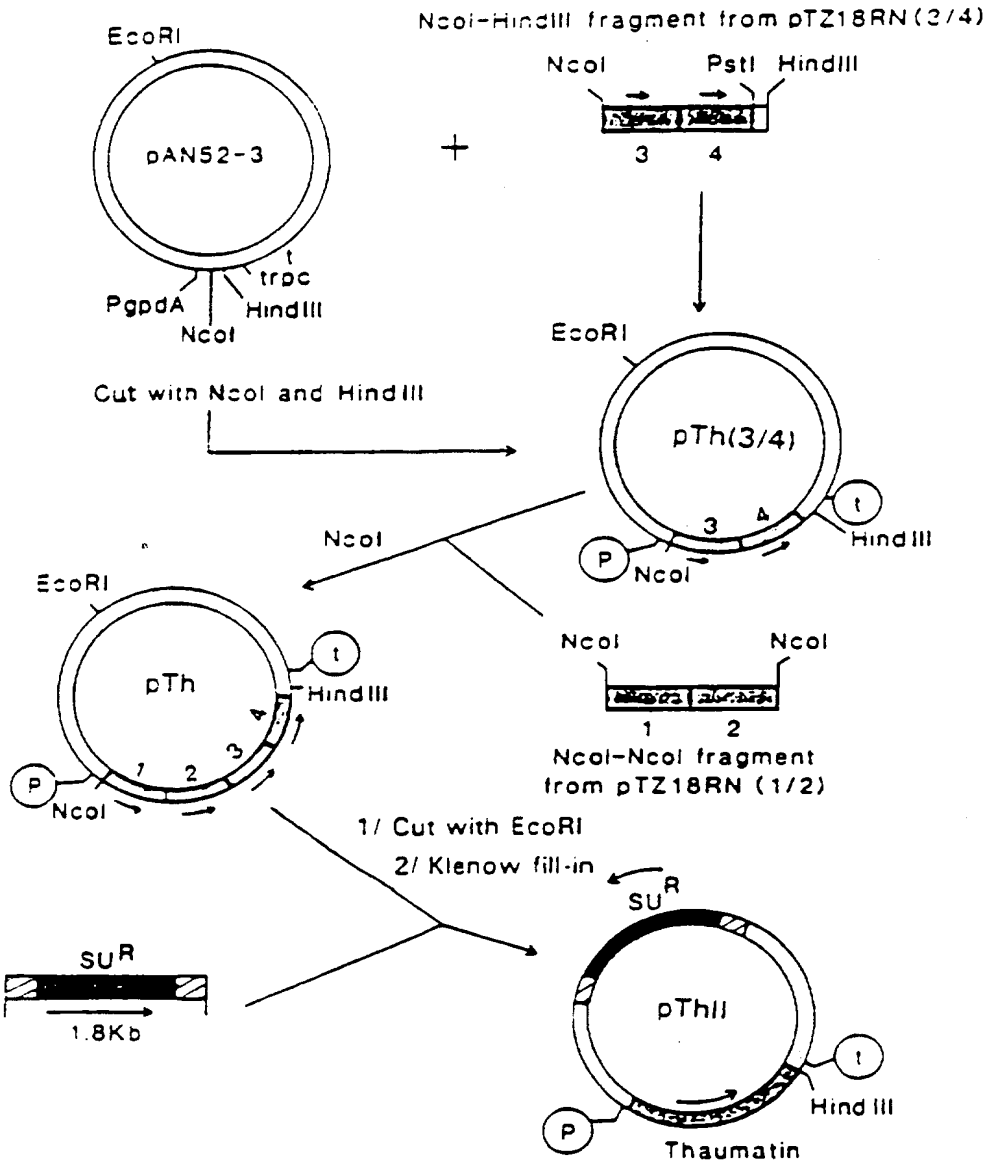


Fig. 6

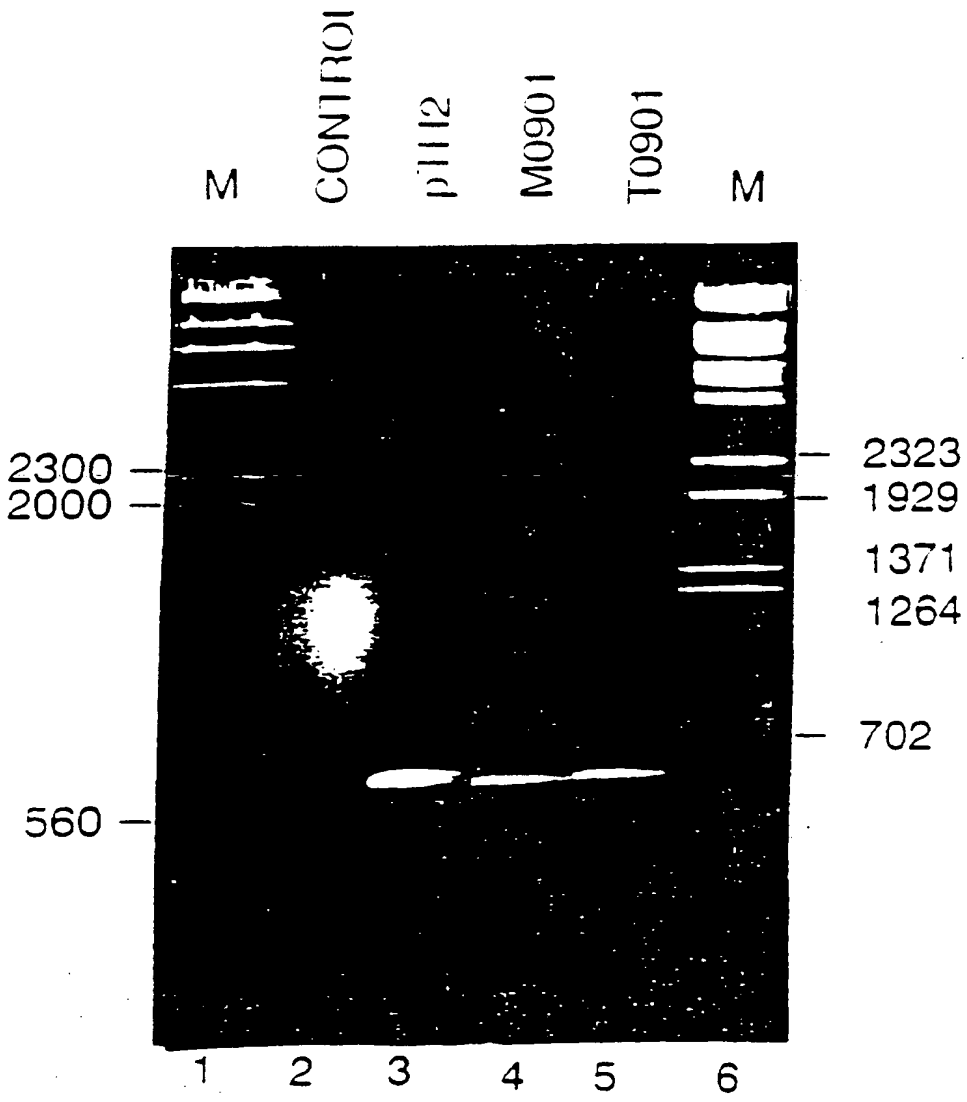


Fig. 7

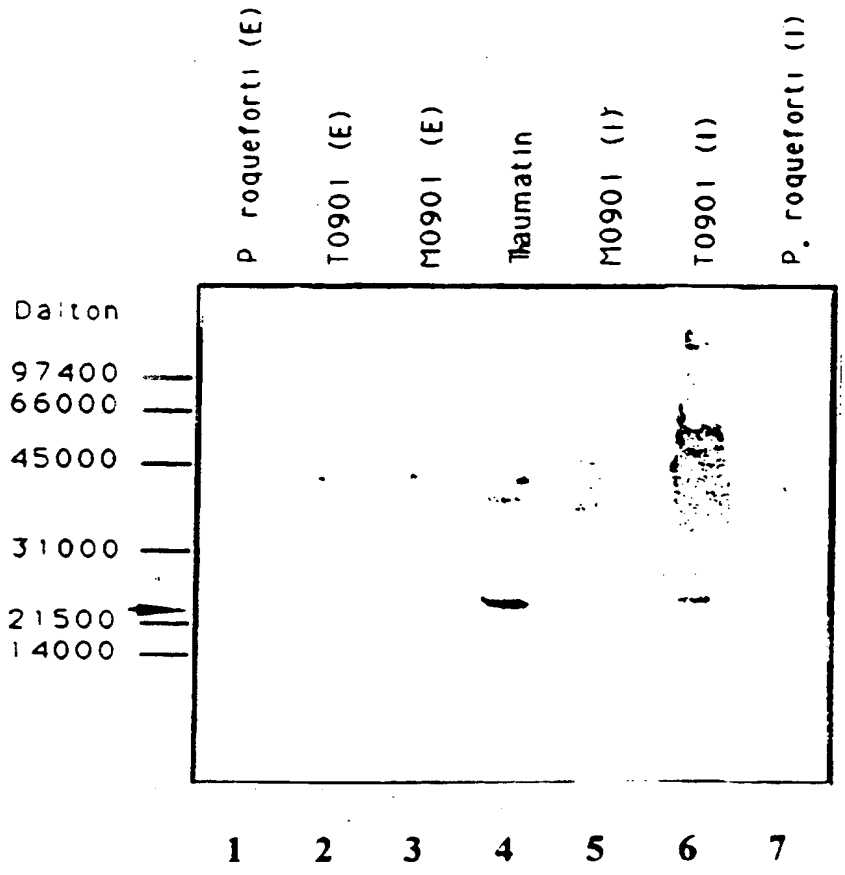


Fig. 8

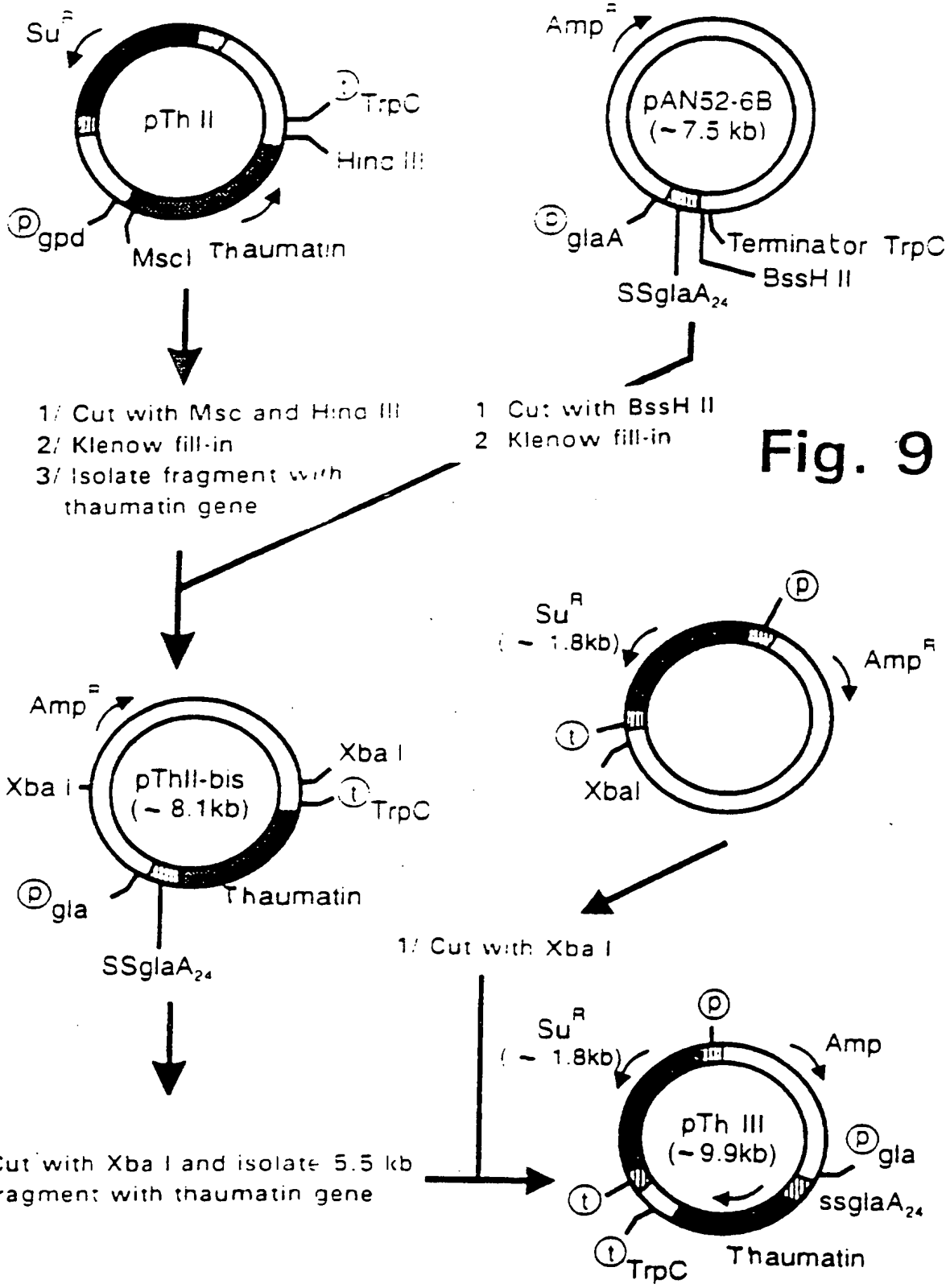


Fig. 9

Fig. 10

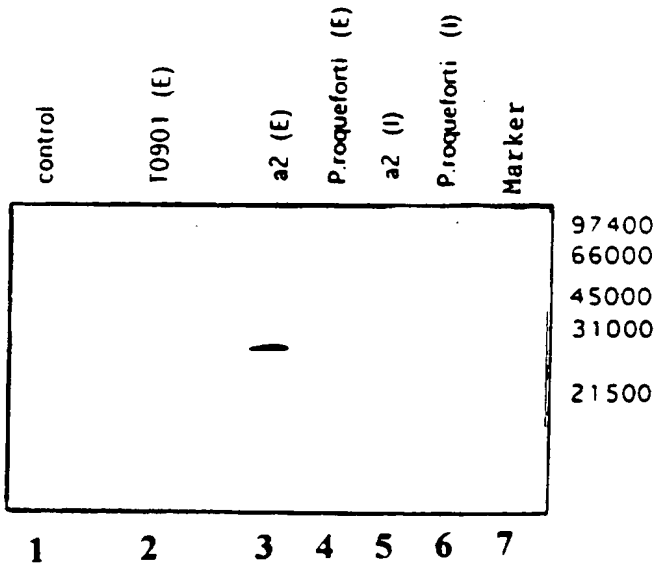


Fig. 11

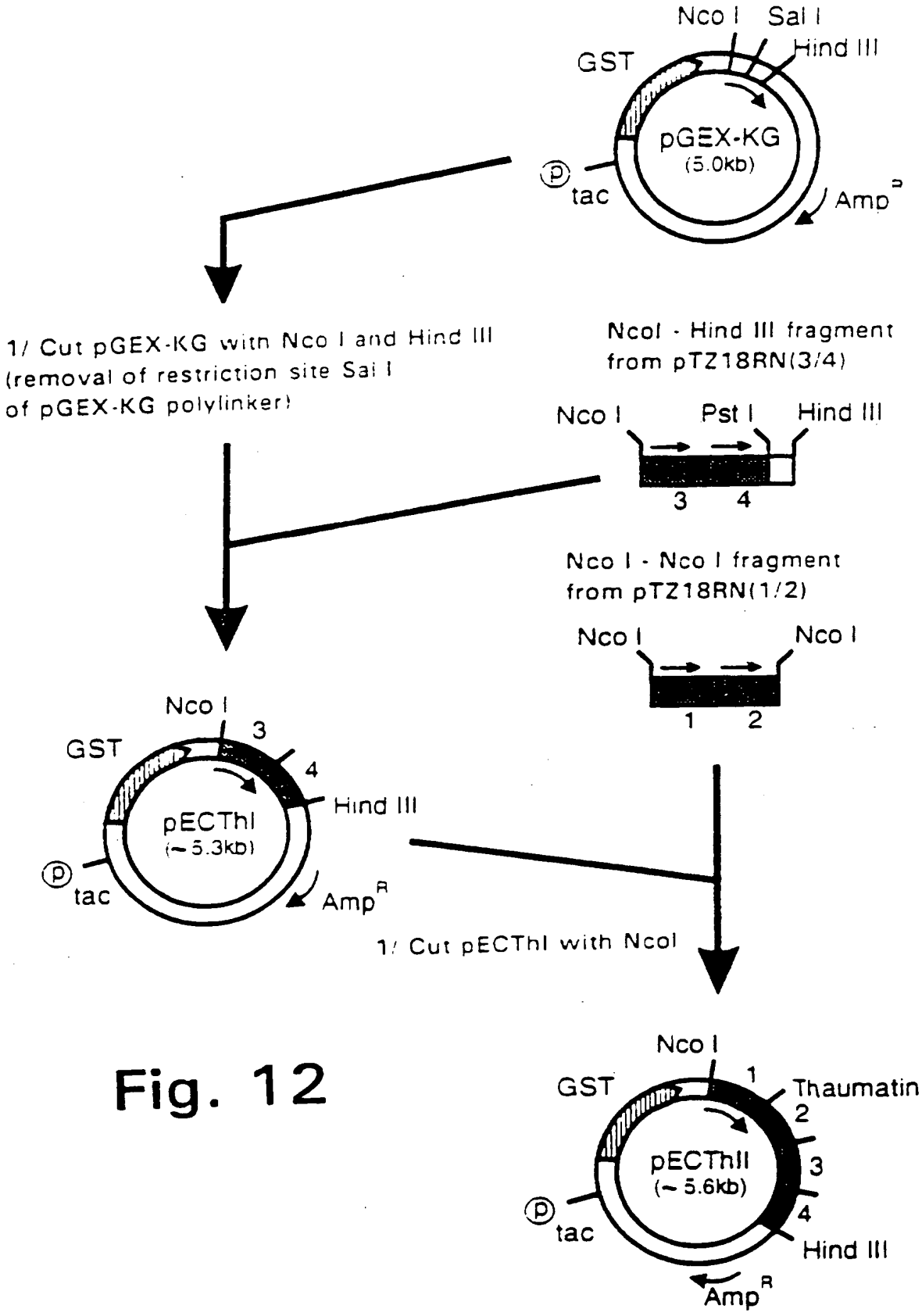


Fig. 12

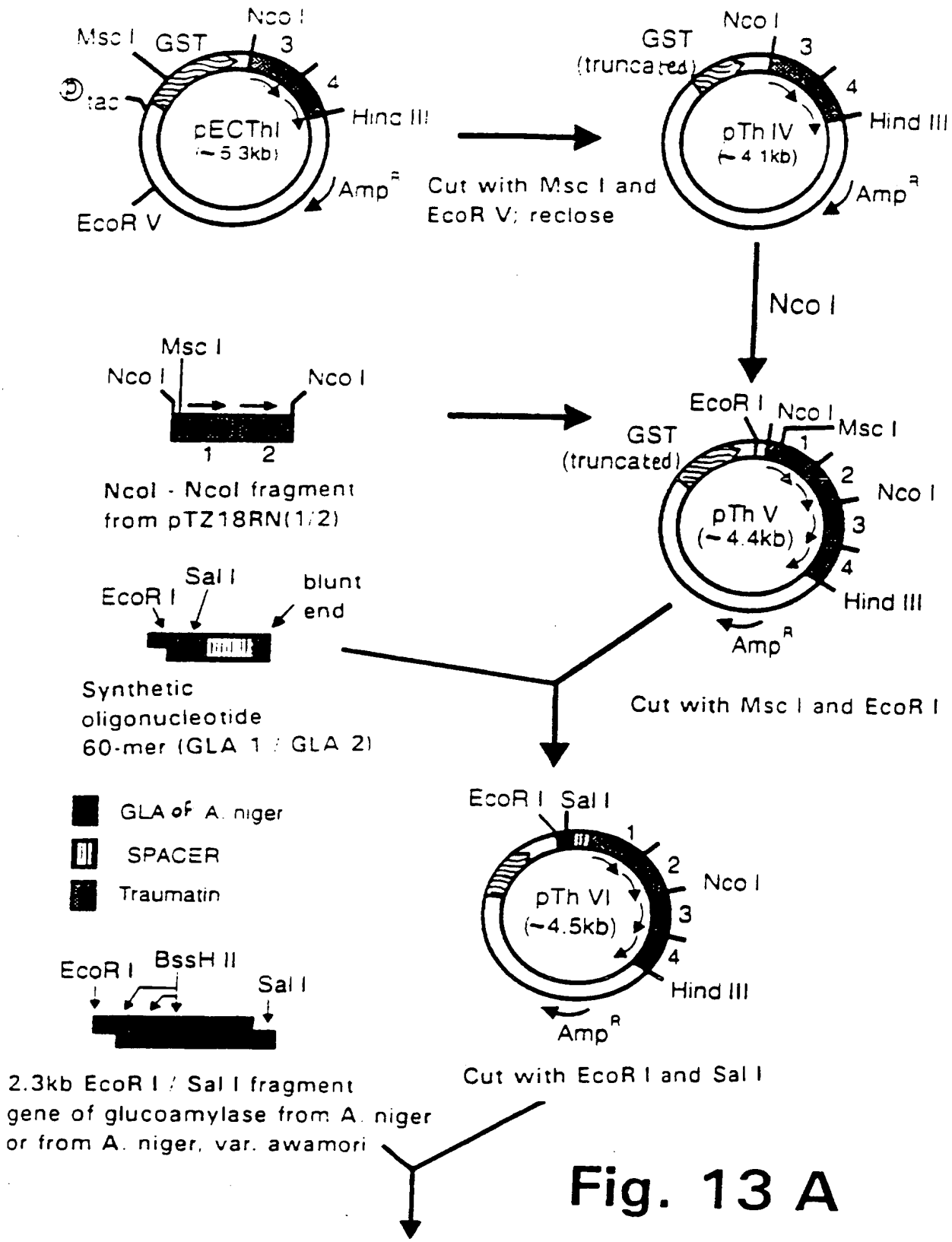


Fig. 13 A

Fig. 13 B

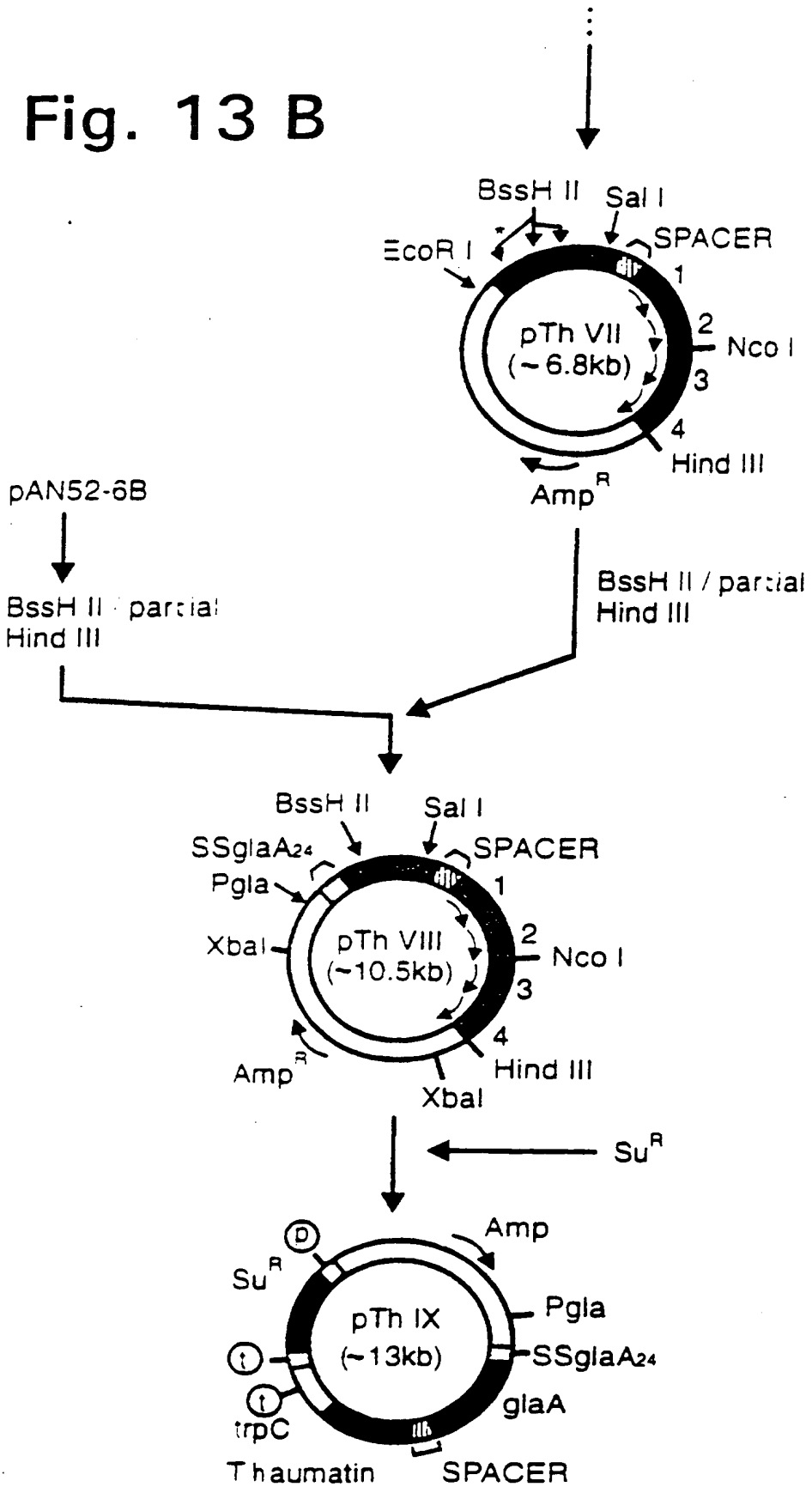
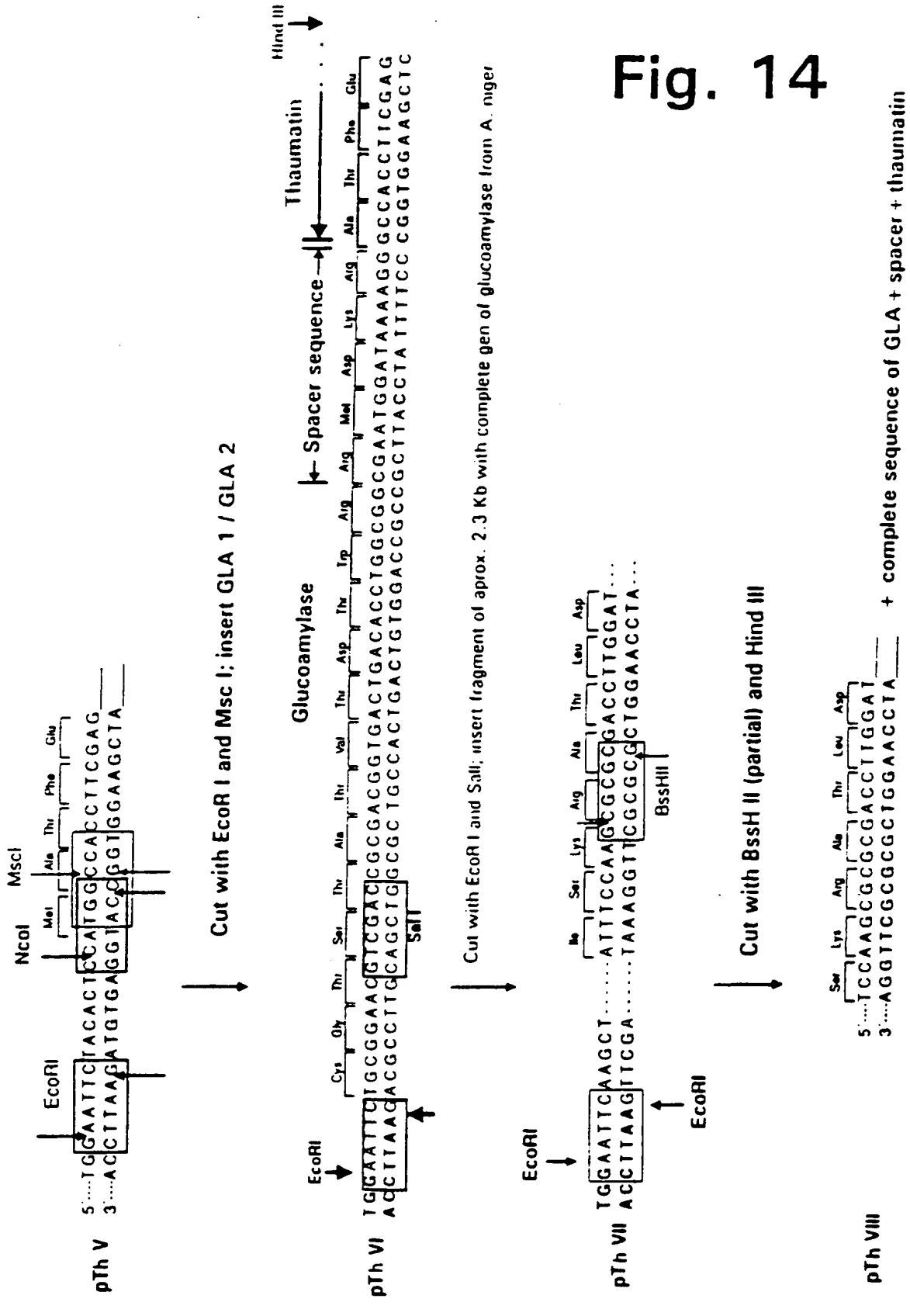


Fig. 14



(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 684 312 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
31.07.1996 Bulletin 1996/31

(43) Date of publication A2:
29.11.1995 Bulletin 1995/48

(21) Application number: **95105973.2**

(22) Date of filing: **21.04.1995**

(51) Int. Cl.[6]: **C12N 15/29**, C07K 14/43, C12N 15/62, C12N 1/15, C12N 15/80 // (C12N1/15, C12R1:66, 1:80)

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **21.04.1994 ES 9400836**

(71) Applicant: **URQUIMA S.A.**
E-08026 Barcelona (ES)

(72) Inventors:
- **Uriach-Marsal, Juan**
  E-08017 Barcelona (ES)
- **Rubio-Susan, Victor**
  E-28008 Madrid (ES)
- **Patino-Martin, Cristina**
  E-28760 Tres Cantos Madrid (ES)
- **Iossif Kalo-Koenova, Eliza**
  E-28028 Madrid (ES)
- **del-Moral-Juarez, Catalina**
  E-28031 Madrid (ES)
- **Faus-Santasusana, Ignacio**
  E-08035 Barcelona (ES)
- **del-Rio-Pericacho, José-Luis**
  E-08224 Tarrasa, Barcelona (ES)
- **Bladé-Piqué, Joan**
  E-08024 Barcelona (ES)

(74) Representative: **Zumstein, Fritz, Dr. et al**
Dr. F. Zumstein
Dipl.-Ing. F. Klingseisen
Bräuhausstrasse 4
D-80331 München (DE)

(54) Preparation process of a natural protein sweetener

(57) Thaumatin II or thaumatin I can be obtained through the expression, not of their natural genes, but of artificial, synthetic and substantially optimized genes following specific rules. Preferably, this expression is carried out in filamentous fungi, especially GRAS fungi and particularly the species Penicillium roquefortii, Aspergillus niger and the awamori variant of Aspergillus niger. Preparing substantially optimized artificial genes for filamentous fungi, performed here for the first time in the case of thaumatin, allows for high protein expression, making the process useful for industrial production of this valuable sweetener. Thaumatins may be obtained extracellularly by using a plasmid with a secretion signal, and also intracellularly. The latter method can be used in animal feed without prior separation from the fungal mycelium.

EP 0 684 312 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 95 10 5973

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
D,Y	AGRIC BIOL CHEM 54 (10). 1990. 2513-2520., XP002004844 HAHM Y T ET AL: "EXPRESSION AND SECRETION OF THAUMATIN FROM ASPERGILLUS-ORYZAE." * the whole document *	1-21	C12N15/29 C07K14/43 C12N15/62 C12N1/15 C12N15/80 //(C12N1/15, C12R1:66,1:80)
Y	BIOCHEMISTRY, vol. 27, 1988, pages 5101-5107, XP002004845 LEE, J.-H., ET AL.: "Expression of synthetic thaumatin genes in yeast" * the whole document *	1-21	
Y	GB-A-2 200 118 (ALLELIX INC) 27 July 1988 * claim 5 *	1-21	
A	MOL GEN GENET 230 (1-2). 1991. 288-294., XP002004846 LLOYD A T ET AL: "CODON USAGE IN ASPERGILLUS-NIDULANS." * page 290 *	1-21	
A	WO-A-87 03007 (BEATRICE CO INC) 21 May 1987 * the whole document *	1-21	TECHNICAL FIELDS SEARCHED (Int.Cl.6) C07K C12N
A	WO-A-90 05775 (INT GENETIC ENG) 31 May 1990 * the whole document *	1-21	
A	EP-A-0 139 501 (BEATRICE CO INC) 2 May 1985 * the whole document *	1-21	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	6 June 1996	Maddox, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)